THE ROLE OF EPIGENETIC FACTORS IN THE PATHOGENESIS OF FAMILIAL X-LINKED MENTAL RETARDATION (XLMR)

Gemma Carvill
BSc (Med) Hons

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

In the Division of Human Genetics, Department of Clinical Laboratory Sciences

UNIVERSITY OF CAPE TOWN

February 2010

DECLARATION

This study was performed from 2006-2010 under the supervision of Dr René Goliath and Professor Jacquie Greenberg and in collaboration with Dr Hans van Bokhoven and Dr Andrew Sharp.

I hereby certify that this is my own unaided work and has not been presented for a degree at any other university

_____

Gemma Carvill
February, 2010

DEDICATION

This thesis is dedicated to my family,

Mum, Dad and my lil bro

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to the following people who have guided me through the last few years:

My supervisor and mentor, Dr René Goliath, for her unwavering support throughout my postgraduate endeavours. Thank you for giving me the freedom to explore my own ideas, for your positive criticisms, enthusiasm and leadership, essentially for being everything a mentor should be, thank you.

Dr Andy Sharp, for his assistance on the DNA methylation aspect of this dissertation. Also, for all the stimulating conversations and for always being willing to share your extensive knowledge and ideas, listening and critiquing my ideas and for your constant support.

Dr Hans van Bokhoven and Dr Arjan de Brouwer, for their assistance on the CNV aspect of this dissertation. Also, for all the support, mentoring and enthusiasm which you bestowed upon me. Finally, a big thank you to everyone in the Nijmegen Department of Human Genetics, as well as all my friends in Nijmegen for all the fun in the lab and out, and for making my stay in The Netherlands so enjoyable.

Prof. Jacquie Greenberg, for your 'open door' policy, many thanks for all the ways in which you have contributed to my development as a researcher, as well as for your unwavering care and concern.

Dr Karen Fieggen and Dr Mike Urban, for all your assistance on all things clinical. Sister Di Sklar for all your help in communicating with the MR families.

Dr Robea Ballo, for her useful comments pertaining to the writing of this dissertation.

My friends in the Division, Janine and Lisa, for always being there, allowing me to natter on about all things science, and for all your valuable insights. To everyone else in the lab, for making this such a wonderful place to work. Finally, thank you to Prof Ramesar and the rest of the HumGen family, who made this a pleasant and pleasurable experience.

To my friends, Jo, Paola, Chris, Liv, Megs (too long a list to mention by name) and the rest of the MAC clan (where many a scientific problem has been solved), thank you for all your support and friendship, there's no way this would of been as much fun without you guys! There's also no way I would be this sane without you all!

The MR families who formed part of this research study, especially family XMR2, you're willingness to do whatever you could to help us is greatly appreciated.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	agenesis of the corpus callosum
AD	autosomal dominant
ADD	ATRX-DNMT3-DNMT3L
ARID	AT-rich domain-interacting domain
ARMR	autosomal recessive mental retardation
ARMS-PCR	amplification refractory mutation system-PCR.
ARX	Aristaless related homeobox
AS	Angelman syndrome
ATP	adenosine triphosphate
ATF7IP1	activating transcription factor 7 interacting protein 2
ATRX	Alpha-thalasemmia mental retardation protien
ATR-X syndrome	X-linked alpha-thalassaemia mental retardation syndrome
BAC	bacterial artificial chromosome
BWS	Beckwith-Wiedemann syndrome
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
CNAG	copy number analysis for genechip
CNP	copy number polymorphism
CNS	central nervous system
CNV	copy number variation
CREB1	cAMP responsive element binding protein 1
DAXX	death-domain associated pro
DD	developmental delay
DECIPHER	database of chromosomal imbalance and phenotype using Ensembl resources
DGV	database of genomic variation
dHPLC	denaturing high performance liquid chromatography
DMR	differentially methylated region
DNA	deoxyribonucleic acid
DNMT	DNA methyltransferase
dNTPS	deoxynucleotide triphosphates
EIEE1	early infantile epileptic encephalopathy-1
ENCODE	encyclopaedia of DNA elements
ESC	embryonic stem cells
ESE	exonic splice enhancer
EtBr	ethidium bromide
EZH2	enhancer of zeste homolog 2
FAS	foetal alcohol syndrome
FDR	false discovery rate
FISH	fluorescent in situ hybridisation
FMR1	fragile X mental retardation protein

FXS	fragile X syndrome
GABA	γ-aminobutyri
HAT	histone acetyltransferase
HbH	haemoglobin H
HDAC	histone deacetyltransferase
HMM	hidden markov model
HMT	histone methyltransferase
HP1	heterochromatic protein
IAP	intracisternal A particle
ICF syndrome	immunodeficiency, centromeric region instability and facial anomalies syndrome
ICR	imprinting control region
IQ	intelligence quotient
ISSX	X-linked infantile spasms syndrome
JARID	Jumonji AT-rich interactive domain
Jmj	Jumonji
KDM5C	lysine (K)- specific demethylase 5C
LCR	low copy repeat
LINE	long interspersed nuclear element
LOD	logarithm of odds
LTD	long term depression
MAP	mitogen-activated protein
MBD	methylated DNA binding domain
MCA	multiple congenital anomalies
MECP2	methyl CpG binding protein 2
MeDIP	methylation dependant immuno-precipitation
microRNA	miRNA
min	minute
MLPA	multiplex ligation-dependant probe amplification
MR	mental retardation
mRNA	messenger RNA
MRX	non-syndromic/non-specific mental retardation
MRXS	syndromic mental retardation
NAHR	non-allelic homologous recombination
NAS	nonsense-associated alternative splicing
NCBI	National Centre for Biotechnology Information
ncRNA	non-coding RNA
NHLS	National Health Laboratory Sciences
NLS	nuclear localisation signal
NMD	nonsense-mediated decay
NRSE	neuron-restrictive silencing element
NSC	neural stem cells
NS-XLMR	non-syndromic/non-specific mental retardation

OMIM	online mendelian inheritance in man
ORD	other repression domain
ORF	open reading frame
PcG	polycomb group proteins
PCR	polymerase chain reaction
PHD	plant homeodomain
PML-NB	promyelocytic leukemia nuclear body
PWS	Prader-Willi syndrome
qPCR	quantitative real-time PCR
RAS	renin-angiotensin system
RE	restriction enzyme
RE1	restriction element 1
REST	restriction element 1 silencing transcription factor
RNA	ribonucleic acid
s	second
SD	segmental duplication
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
S-XLMR	syndromic mental retardation
TCA	Trichostatin A
TtxG	trithorax group proteins
UCE	ultraconserved elements
UCT	University of Cape Town
UTR	untranslated region
VNTR	variable number of tandem repeats
WHS	Wolf-Hirschhorn syndrome
XLAG	X-Linked lissencephaly with ambiguous genitalia
XLMESID	X-linked myoclonic epilepsy with spasticity and intellectual disability
XLMR	X-linked mental retardation

ABSTRACT

Mental retardation (MR) is a handicap with severe implications not only for those that suffer from this disability, but also for their families, society and the welfare systems which support them. A large proportion of these individuals are afflicted with the X-linked form of the condition. To date a total of 87 genes have been implicated in the pathogenesis of X-linked mental retardation (XLMR). A number of these XLMR genes (~22%) are known or putative transcriptional regulators, which control gene regulation via various epigenetic mechanisms. Integral to the functioning of these epigenetic mechanisms is DNA methylation, occurring at CpG islands located in gene promoters where methylation is associated with gene silencing. In this study it was hypothesised that alterations to the DNA methylation profile were a molecular feature of MR in patients positive for mutations in putative 'epigenetic' XLMR genes.

In order to test this hypothesis it was first necessary to identify patients with disease-causing XLMR mutations. Given the complex genetic landscape of MR, several molecular approaches were employed in order to achieve this objective. Analysis of genomic copy number in a MR patient cohort using microarray technologies revealed three disease-causing mutations. These included a 17p11.2Del in an isolated case of MR which results in the Smith Magenis syndrome in this patient. Also, a de novo Xq25Dup was detected in an isolated male MR case. Finally, a large X-chromosomal deletion (Xp26.3-27.3Del) was identified in a familial case of XLMR; this aberration exhibits low penetrance in the family. X-chromosomal linkage analysis in four XLMR families defined a critical interval in two of these families. A disease-causing mutation in the *ATRX* gene was subsequently identified by positional candidate gene screening in one family. In the second family with a defined disease-associated chromosomal interval, no mutations were identified by positional candidate screening of known XLMR genes in the region. This result suggested that this family may harbour a mutation in a novel XLMR candidate gene. Finally, functional candidate gene screening in two commonly mutated XLMR genes (*ARX* and *KDM5C*) led to the identification of two disease-causing mutations in the *ARX* gene (c.428_451dup24) and a putative disease-causing mutation (c.2517-

7_8insTAC) in *KDM5C*. The outcomes of the mutation screening phase of this dissertation have resulted in the addition of two diagnostic tests (*ARX* and *ATRX*) to the repertoire of tests offered to patients with MR in South Africa. Also, the diagnostic protocol for copy number analysis using microarrays is now in the development stage.

The *ATRX* mutation positive family identified in this study was shown to exhibit altered DNA methylation at *ATF7IP2* by whole-genome DNA methylation profiling. This methylation profiling entailed methylation dependant immunoprecipitation (MeDIP) and subsequent hybridisation to a whole-genome CpG island and promoter array using DNA from four affected and three unaffected males. The methylation fold changes generated for each individual were used to identify differentially methylated regions (DMRs) between patient and control samples using a novel statistical approach. From these analyses, the *ATF7IP2* promoter region exhibited significant hypermethylation in the *ATRX* mutation positive individuals. This gene, *ATF7IP2*, has been shown to interact at a protein level with MBD1; known to play a role in the formation of silent chromatin complexes, much like ATRX. The nature of this interaction and its putative ramifications was hypothesised.

The findings presented in this dissertation therefore suggested that transcriptional regulator XLMR genes may affect the DNA methylation profiles of mutation positive MR patients in a limited manner, at least for *ATRX* patients. Thus, the findings of this study, as well as the methodology presented, has provided a novel research strategy which can be employed in order to elucidate the role of this critical epigenetic mechanism in the pathogenesis of XLMR/MR.

CHAPTER 1

LITERATURE REVIEW

Mental retardation (MR) is estimated to have a prevalence of 1-3% in the developed world, making it a common congenital disorder (Leonard and Wen. 2002). Patients afflicted with MR have a decreased ability to adapt their behaviour to the changing environment, rendering them dependent on constant support from families and/or social workers, as well as the country's welfare systems. Furthermore, patients often require specialist medical treatment owing to the high comorbidity rate of MR with additional clinical features/disorders. The characteristics and implications of this handicap make MR an important medical and socio-economic-impacting disorder.

In developing countries, there tends to be an increased incidence of MR, with factors such as malnutrition, cultural deprivation and poor healthcare hypothesised to contribute to this increase (Durkin. 2002). This enhanced MR incidence further impacts on the already underfunded and overstretched welfare and public health systems, particularly in developing countries such as South Africa.

MR is one of the most common reasons for referral to paediatric, neurological or genetic clinics. Despite this, MR often cannot be ascribed to a specific genetic or environmental cause, therefore limiting the prognosis and monitoring of a patient's condition, as well as accurate genetic counselling of the family.

Finally, the medical and socio-economic impact of the condition is exascerbated by the lack of treatment for almost all forms of MR. This lack of available treatments reflects our limited understanding of MR pathogenesis and highlights the importance of MR research aimed at characterising patients' deficits in cognitive development and functioning at a molecular level.

1.1 MENTAL RETARDATION

MR is a complex and clinically heterogeneous condition which lacks an entirely comprehensive description. The most commonly used definition is as described by the American Psychiatric Association, which reads: 'This disorder is characterised by significantly subaverage intellectual functioning (an intelligence quotient (IQ) of approximately 70 or below) with onset before 18 years and concurrent deficits or impairments in adaptive functioning (American Psychiatric Association. Task Force on DSM-IV. 1994). Moreover, MR can be characterised according to the degree of severity of the condition, as measured by the IQ score (Table 1.1). MR can be further classified according to clinical presentation into two groups: syndromic MR in which the cognitive deficits present in combination with additional clinical features and non-syndromic features in which MR is the only clinical feature (Mulley et al. 1992).

Table 1.1 Classes of MR as defined by IQ score (American Psychiatric Association. Task Force on DSM-IV. 1994).

Class of MR	**IQ level range**
Mild	50-55 to ~70
Moderate	35-40 to 50-55
Severe	20-25 to 35-40
Profound	Below 20-25

In developed countries the prevalence of mild MR is estimated to be 1-3%, while more severe forms (IQ<50) occur in only 0.3-0.5% of the population, as reviewed previously (Stevenson, Schwartz and Schroer. 2000). In South Africa, studies focused on estimating the prevalence of the disorder are limited. However, a study by Christianson and colleagues (2002) showed an MR prevalence of 3.6% in a rural community, suggesting local estimates may reflect those in the international community (Christianson et al. 2002).

The causes of MR are diverse and can be attributed to a myriad of both genetic and environmental factors. The most common pathogenic environmental factors include perinatal asphyxia, neonatal or postnatal central nervous system (CNS) infection and intrauterine exposure to toxins (including alcohol exposure) (Vasconcelos. 2004). In South Africa and particularly the Western Cape, foetal alcohol syndrome (FAS) is one of the leading contributors to the burden of MR, with a prevalence of up to 67.2 - 89.2 cases per 1000 children (May et al. 2007, Urban et al. 2008).

It has been estimated that 60-75% of severe MR cases and 38-55% of mild cases can be attributed to a specific cause (Inlow and Restifo. 2004). Of these diagnoses it is estimated that genetic causes account for approximately 25-50% of severe MR (McLaren and Bryson. 1987). In comparison, mild MR can be attributed to a genetic cause in a much smaller percentage (5-17%) of cases (Aicardi. 1998, King, Rotter and Motulsky. 2002).

The exact contribution of genetics to the pathogenesis of MR remains unknown, difficult to estimate, and exhibits inter-study variability. However, it remains evident that there is a strong hereditary component to the pathogenesis of MR, thus exemplifying the necessity for genetic MR research in order to enhance clinical management of the disorder. With this in mind, this study embraced a two tiered approach. Firstly, in order to address the lack of molecular MR diagnoses, particularly in a developing country such as South Africa, a variety of molecular techniques were employed in order to identify underlying disease-causing alterations in a group of South African patients. The results of these investigations would guide the expansion of the repertoire of MR diagnostic tests offered to South African MR patients. Secondly, in order to contribute to the understanding of MR pathogenesis, the downstream effects of disease-causing alterations on epigenetic mechanisms was assessed. More specifically, potential alterations to the DNA methylation profile in mutation positive individuals were investigated.

1.2 GENETIC ETIOLOGY OF MR

The genetic causes of MR are diverse and can be roughly distinguished into three primary groups. The first comprises copy number changes (chromosomal aneuploidies and gross/minor chromosomal aberrations) which generally exert their pathogenic effects by altered gene dosage. The second group encompasses the imprinting disorders which manifest due to disruptions at loci controlled by monoallelic gene expression. Lastly, the third group consists of monogenic disorders arising from mutations in single genes leading to anomalous protein function (Chelly et al. 2006).

In general, gene dosage effects and imprinting disorders are associated with the syndromic form of MR, while non-syndromic forms of the disorder occur primarily in monogenic forms of MR. However, this over-simplification is increasingly being challenged with the identification of patients with small copy number variations (CNVs) having a non-syndromic clinical presentation (Lugtenberg et al. 2006). In addition, a number of monogenic disorders, specifically X-linked MR (XLMR), are associated with a clinical presentation reflecting both forms of the disorder, even in patients with the same mutations (Chiurazzi, Tabolacci and Neri. 2004).

1.2.1 CHROMOSOMAL ABERRATIONS

Chromosomal aneuploidies are the largest of the genomic abnormalities and are generally the result of errors during meiosis (Table 1.2). Chromosomal monosomy is rare, with the karyotype 46 XO (Turner syndrome), the only monosomy compatible with life. Patients with Turner syndrome often present with learning disabilities particularly in spatial reasoning and mathematics. However, chromosomal trisomies are more commonly seen in pregnancies that proceed to term and include Edwards syndrome (trisomy 18) and Patua syndrome (trisomy 13). However, individuals born with these disorders generally do not survive beyond 18 months of age owing to the severity of these congenital abnormalities. Finally, Down syndrome (trisomy 21) is the only chromosomal aneuploidy in which patients survive till adulthood. Down syndrome has a prevalence of 1 in 800 live births making it one of the most common genetic causes of MR.

Besides whole chromosomal aneuplodies, it has also been established that partial chromosomal copy number changes or translocations can cause MR (Table 1.2). Traditionally these chromosomal aberrations have been detected using conventional cytogenetic techniques such as G-banded karyotyping. However, this technique has a limited resolution of 5-10Mb and therefore allows chromosomal aberration detection in only 3-5% of patients with MR (Stankiewicz and Beaudet. 2007).

Molecular cytogenetic techniques such as FISH (fluorescent in situ hybridisation) and molecular methods including MLPA (multiplex ligation-dependant probe amplification) have been developed with the aim of increasing the ability to detect smaller copy number changes. These techniques have been successfully developed to investigate genomic regions harbouring known microdeletion syndromes (Table 1.2). In addition, these molecular techniques can be adapted to target subtelomeric regions which are prone to re-arrangement and account for 5-6% of all MR (Flint and Knight. 2003).

With the advent of microarray technologies and the application of these platforms to detect chromosomal changes, the ability to detect even smaller copy number changes has become a reality. These smaller copy number changes are termed 'Copy Number Variations (CNVs)' and are defined as: DNA segments greater than 1 kb in length that are present in fewer or more copies than expected in the genome that have not arisen from the insertion or deletion of transposable elements (Zahir and Friedman. 2007). Numerous international studies have employed array technology in order to identify MR-causing CNVs. These studies report a variable pick-up rate of between 4-24% (Aston et al. 2008, de Vries et al. 2005, Fan et al. 2007, Friedman et al. 2006, Menten et al. 2006, Rosenberg et al. 2005, Rosenberg et al. 2006, Schoumans et al. 2005, Shaw-Smith et al. 2004, Slater et al. 2005, Tyson et al. 2005, Vissers et al. 2003). The variability observed between detection rates is owing to differential clinical inclusion and exclusion criteria, as well as the microarray design employed. The highest pick-up rates were observed when high resolution tiling arrays were used and MR patient cohorts consisted of patients in whom additional dysmorphic and clinical features were present (Rosenberg et al. 2006, Shaw-Smith et al. 2004). The role of CNVs in the pathogenesis of MR is discussed in further detail in Chapter 2 of this dissertation.

Table 1.2 Examples of chromosomal copy number changes and the molecular and cytogenetic techniques employed to detect them.

Type of abnormality	Examples	Technique
Chromosomal aneuploidies Partial deletions Partial duplications Translocations	Down syndrome Turner syndrome Cri de Chat syndrome (5p) 4p, 9q (Chelly et al. 2006)	Cytogenetic techniques
Contiguous gene syndromes Subtelomeric aberrations	Wolf-Hirschhorn syndrome Williams syndrome Rubenstein Taybi syndrome Smith Magenis syndrome DiGeorge syndrome 22q11 deletion syndrome (Chelly et al. 2006)	Molecular cytogenetic (FISH) and molecular (MLPA) techniques.
Genome-wide duplication and deletions	Several new microdeletion/duplication syndromes detected using array technologies e.g. 15q24 microdeletion syndrome (Sharp et al. 2007) 17q21.31 microdeletion syndrome (Koolen et al. 2006)	Array technologies

1.2.2 IMPRINTING DISORDERS

Genomic imprinting involves monoallelic expression of genes at certain genomic loci in a parent-of-origin dependent manner. This monoallelic expression is achieved through the imprinting control region (ICR) which establishes DNA methylation patterns in a parent-of-origin manner. Of the approximately 60 imprinted loci known to date (catalogued at http://www.geneimprint.com) a large proportion of these genes are expressed in the brain in both a spatial and temporal fashion (Wilkinson, Davies and Isles. 2007).

Perhaps the most well-known of these neurodevelopmentally-associated imprinted loci is located at 15q11.2-q13 (Figure 1.1). Paternally derived deletions at this locus cause Prader-Willi syndrome (PWS) occurring with a frequency of 1 in 20000 births. PWS is characterised by MR, failure to thrive in infancy, hyperphagia and obesity in childhood and behavioural problems (Goldstone. 2004). Conversely, maternally derived deletions within this region result in Angelman syndrome (AS), occurring with a frequency of 1 in 15000 births and characterised by MR, speech impairment and behavioural problems (Lossie et al. 2001).

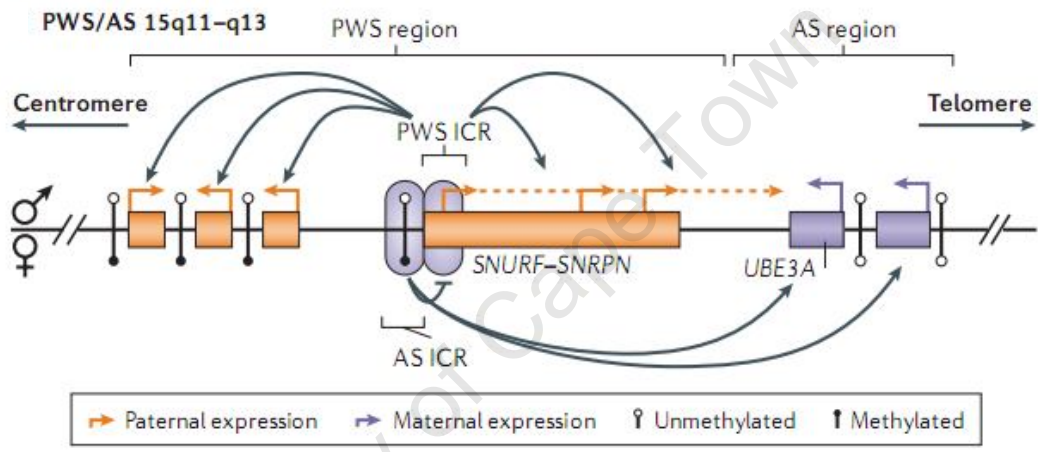


Figure 1.1 The 15q11.2-q13 imprinted locus associated with Prader-Willi/Angelman syndrome (PWS/AS). This locus contains two imprinting control regions (ICR) which drive the parentally-determined monoallelic expression of genes in the region (as indicated by the black arrows). This allele-specific expression as determined by the ICRs is maintained by DNA methylation and additional histone modifications. Paternally derived deletions in the region lead to PWS, as gene expression cannot occur from the imprinted maternal locus, conversely AS occurs due to deletions of the maternally derived allele (adapted from Bienvenu et al. 2006)

While these two disorders are defined by distinct phenotypes, both exhibit MR as a defining feature, thus illustrating the importance of these imprinted loci in neuronal development. Furthermore, a second imprinted locus (8q24) has recently been implicated in MR, with mutations in the maternally-expressed *KCNK9* leading to Birk Barel MR dysmorphism syndrome (Barel et al. 2008). To date, 15q11.2-q13 and 8q24 are the only imprinted loci to be associated with MR, however, given the proportion of imprinted genes expressed in the brain, imprinted loci are attractive candidates for MR pathogenesis.

1.2.3 MONOGENIC DISORDERS

The number of monogenic causes of MR are biased towards the X-linked form of the disorder with over 80 XLMR genes identified to date (Chiurazzi et al. 2008). By comparison, the exact number of autosomal MR genes is difficult to quantify given the clinical and genetic heterogeneity of the disorder. In 2004 an extensive literature review and text mining approach using (amongst other resources) the online mendelian inheritance in man (OMIM) database at NCBI estimated that the number of MR genes stands at 282 (including XLMR genes) (Inlow and Restifo. 2004). The original dataset used for this study was obtained from OMIM using the search phrase 'mental retardation' which returned a total of 1010 entries (Inlow and Restifo. 2004). By comparison, in 2009 (December) a search for the same phrase in OMIM returned a total of 1626 'hits' clearly illustrating the increasing number of MR genes (and CNVs). This figure is most likely due to recent technological advances in microarray applications for CNV detection, homozygosity mapping and diagnostics as well as improvements in high-throughput DNA sequencing techniques. These technologies allow a more rapid and in-depth approach to molecular MR diagnoses. Furthermore, these additional OMIM entries reflect an increased number of entries for a single MR locus, due to differential clinical diagnosis. These observations furthermore illustrate the extent of both the clinical and genetic heterogeneity of MR.

While the exact number of MR genes remains unknown, there are a number of well known autosomal MR monogenic genes. Mutations in these genes primarily exhibit an autosomal dominant (AD) inheritance pattern and include a variety of syndromes such as: Rubenstein-Taybi syndrome (*CBP* mutations), DiGeorge and velocardiofacial syndrome (*TBX1* mutations, as well as larger chromosomal deletions) and Smith-Magenis syndrome (mutations in *RA1*, but also loci-specific deletions) (Petrij et al. 1995, Slager et al. 2003, Yagi et al. 2003). These ADMR genes (and others) have principally been identified through positional candidate gene mutation screening at either translocation breakpoints or at common microdeletion loci where patients have tested negative by conventional deletion screening. However, the rare occurrence of patients with these types of mutations has hindered the identification of additional AD genes as has the common de novo nature of mutations which prohibits linkage analysis.

Until recently the vast majority of autosomal recessive mental retardation (ARMR) genes remained elusive. However, with the advent of DNA microarray technologies it has now become possible to conduct whole-genome homozygosity mapping in consanguineous families. Recently, using this approach, several ARMR genes have been identified, they include: *PRSS12*, *CRBN*, *CC2DIA*, *GRIK2* and *TUSC3*, *TRAPPC9, FA2H* (Basel-Vanagaite et al. 2006, Higgins et al. 2004, Mir et al. 2009, Molinari et al. 2002, Motazacker et al. 2007, Dick et al. 2010). It is anticipated that the application of this technique will lead to the identification of a substantial proportion of ARMR genes.

1.2.3.1 X-Linked Mental Retardation (XLMR)

It has been known for over 30 years that MR is more prevalent in males than females. This preponderance is attributed, in part, to mutations in X chromosomal genes (X-linked Mental Retardation – XLMR) which are estimated to account for 10-15% of all MR (Ropers and Hamel. 2005). To date there have been over 80 XLMR genes implicated in MR, a figure disproportionally higher than autosomal monogenic causes. This preponderance of XLMR genes led to the hypothesis that genes involved in cognitive development cluster on the X chromosome, possibly due to positive sexual selection (Zechner et al. 2001). Opponents to this theory suggested that the increased number of XLMR genes was merely due to an ascertainment bias owing to the relative ease with which X-linked traits can be mapped in hemizygous males as well as the fortuitous availability of large pedigrees. However, support for this hypothesis was outlined by Inlow and Restifo (2004) where these authors concluded that there is a 1.9 fold overrepresentation of MR genes on the X chromosome and that this observation was not due to ascertainment bias (Inlow and Restifo. 2004). This empirical evidence suggests that those genes involved in cognitive development and functions are overrepresented on the X chromosome.

XLMR phenotypes

As aforementioned, MR can be categorised into two clinically distinct groups, those that manifest with additional clinical phenotypes (syndromic) and those that present with MR only (non-syndromic). With specific reference to XLMR these groups are designated non-syndromic/non-specific mental retardation (MRX or NS-XLMR) and

syndromic mental retardation (MRXS or S-XLMR) (Chiurazzi, Tabolacci and Neri. 2004, Kleefstra and Hamel. 2005). Furthermore, the MRXS class can be further sub-divided into two groups (Chiurazzi et al. 2008):

1. Syndromes: characterised by multiple congenital anomalies (MCA)
2. Neuromuscular disorders: characterised by muscular or neurological abnormalities (epilepsy, seizures, dystonia, spasticity, muscle weakness etc).

To date 215 conditions have been described, comprising 98 syndromes, 51 neuromuscular phenotypes and 66 MRX conditions (Chiurazzi et al. 2008).

XLMR genes

Of the 215 XLMR conditions described previously, 97 have been mapped to definitive loci and 82 genes have been cloned (Chiurazzi et al. 2008). Furthermore, over the past year, five additional XLMR genes (*CASK*, *SYP*, *ZNF711*, *SIZN1*, *IAP*) have been reported, bringing the total to 87 X-linked genes involved in the pathogenesis of MR (Figure 1.2) (Cho et al. 2008, Molinari et al. 2008, Tarpey et al. 2009).

These genes have been identified using a number of strategic molecular approaches. Firstly, a large collaborative study conducted at the Sanger Institute, Cambridge, UK resulted in the identification of the XLMR genes *CASK*, *SYP* and *ZNF711* (as well as six others included in the review by (Chiurazzi et al. 2008) by re-sequencing of 718 Vega-annotated X-chromosomal genes (at a coverage of 75%) in 208 XLMR family probands (Tarpey et al. 2009). Secondly, a functional candidate gene approach was used to identify two novel XLMR genes. The first, *SIZN1* (*ZCCHC12*), was selected for mutation screening in a group of XLMR patients based on its localisation to the X chromosome, basal forebrain cholinergic neuronal expression and SIZN1 function (Cho et al. 2008). The second novel gene, *IAP* (*MAGT1*), is a paralogue of the recently identified autosomal recessive non-syndromic mental retardation gene *TUSC3* and was therefore selected for mutation screening in a XLMR cohort (Molinari et al. 2008). Taken together, these studies illustrate the importance of adopting multiple strategies for the identification of novel XLMR genes.

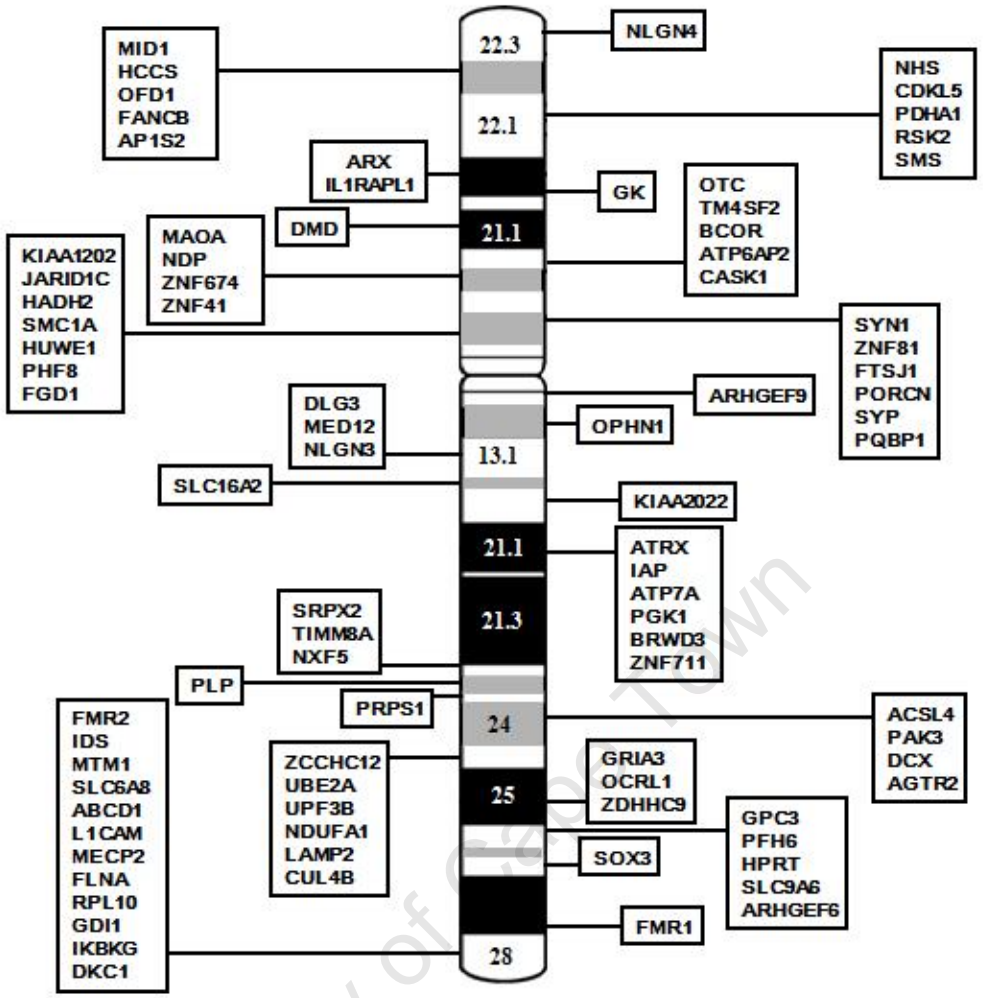


Figure 1.2 Ideogram of the X chromosome depicting the chromosomal location of all 87 known XLMR genes. (Adapted from Chiurazzi et al. 2008).

Despite the considerable number of XLMR genes that have been identified, the majority of XLMR cases cannot be attributed to a molecular cause. A review of approximately 600 XLMR families which comprise the EuroMRX consortium showed that the underlying etiology was known in only 42% of XLMR families (with at least one obligate female carrier) and 17% of brother-pairs (2-5 sibs) (de Brouwer et al. 2007). Even the recent X-chromosomal re-sequencing project, which led to the identification of nine novel XLMR genes has not vastly improved these numbers (Tarpey et al. 2009). In this study, the 208 patients investigated by Tarpey and colleagues (2009) were mutation-negative for cytogenetic abnormalities as well as for unambiguous disease-causing mutations in the known XLMR genes. The nine genes identified in this study account for 25% of families investigated collectively (Tarpey et al. 2009).

Of the known XLMR genes, there is a variable mutation frequency that is gene specific. By far the most common form of XLMR is the fragile X syndrome (FXS), caused by a CGG repeat expansion in the promoter of the *fragile X mental retardation protein* gene (*FMR1*) and accounting for 15-25% of XLMR (Kleefstra and Hamel. 2005). Mutations in the *Aristaless related homeobox* (*ARX*) gene are the second biggest contributor to XLMR. *ARX* mutations account for up to 9.5% of XLMR in patients that are negative for the FXS repeat mutation (Poirier et al. 2006). The etiology of the remaining patients can be attributed to mutations in a number of XLMR associated genes, each accounting for a small proportion of the total incidence. These include; *MECP2, SLC6A8, JARID1C, IL1RAPL1, PAK3, DLG3* and *FTSJ1* (Table 1.3) (de Brouwer et al. 2007).

Table 1.3 Mutation frequencies of XLMR genes in the EuroMRX families (adapted from de Brouwer et al. 2007)

Rank	Gene symbol	Gene name	OMIM accession number	Frequency in XLMR family (%)	Frequency in brother pairs(%)
1.	*ARX*	*Aristaless related homeobox*	***300382**	7.5	1.5
2.	*MECP2*	*Methyl CpG binding protein 2*	***300045**	6.2	0
3.	*OPHN1*	*Oligophrenin 1*	***300127**	4.8	4.6
4.	*PQBP1*	*Polyglutamine binding protein 1*	**+300463**	4.5	0
5.	*KDM5C (JARID1C)*	*Lysine (K)- specific demethylase 5C*	***314690**	4.2	4.3
6.	*ACSL4*	*acyl-CoA synthetase long-chain family member 4*	***300157**	3.1	1.7
7.	*FTSJ1*	*FtsJ homolog 1*	***300499**	2.5	1.2
8.	*SLC6A8*	*solute carrier family 6 (neurotransmitter transporter, creatine), member 8*	***300036**	1.8	1.8
9.	*IL1RAPL1*	*interleukin 1 receptor accessory protein-like 1*	***300206**	1.5	1.7
10.	*PHF8*	*PHD finger protein 8*	***300560**	1.3	0

The high proportion of XLMR families whose phenotype has not been defined by a molecular cause highlights the importance of the application of novel techniques to identify XLMR genes or disease-mechanisms. High-throughput DNA sequencing technologies such as those described by Tarpey et al. (2009) will certainly play a role in future identification of XLMR genes in large XLMR cohorts. Furthermore, the application of advanced DNA sequencing methodologies will allow the detection of

disease-causing regulatory variants (e.g. in promoter/enhancer/silencer regions). While this methodology is likely to enhance the number of disease-causing DNA sequence variants, it is hypothesised that a significant proportion of XLMR will be attributed to, as yet, unknown disease-mechanisms. Candidate alternative pathogenic mechanisms could include X-chromosomal DNA methylation aberrations. Therefore, it is essential that the existing DNA methylation technologies are adapted to identify aberrant DNA methylation in XLMR patients. Finally, it has been hypothesised that the male excess of MR may be attributed to X-linked 'risk factors', genetic variations which predispose to, rather than cause MR (Mandel and Chelly. 2004). This multifactorial hypothesis has been expanded to include variations which may occur on the autosomes and/or include CNVs which constitute a 'general MR predisposition' in males (Raymond and Tarpey. 2006). While an attractive hypothesis, to date, no 'risk factors' of MR have been identified to support this polygenic model.

Molecular pathology of XLMR genes

The 87 known XLMR genes are involved in a host of cellular processes and have a wide variety of functions within these processes (Figure 1.3). Not surprisingly, those genes which play a role in signal transduction account for a substantial proportion of XLMR (18%) as do proteins found in the cell's 'membrane component' (16%) which participate in cell-signalling. Together, these biological processes are vital for neuronal cell-cell communications which contribute to synaptic plasticity and correct neuronal development. However, the largest number of XLMR genes are either known or postulated to play a role in transcription regulation (22%). Transcription regulation is a complex process which is epigenetically controlled so as to induce a chromatin state which is either open (active gene transcription) or closed (gene silenced).

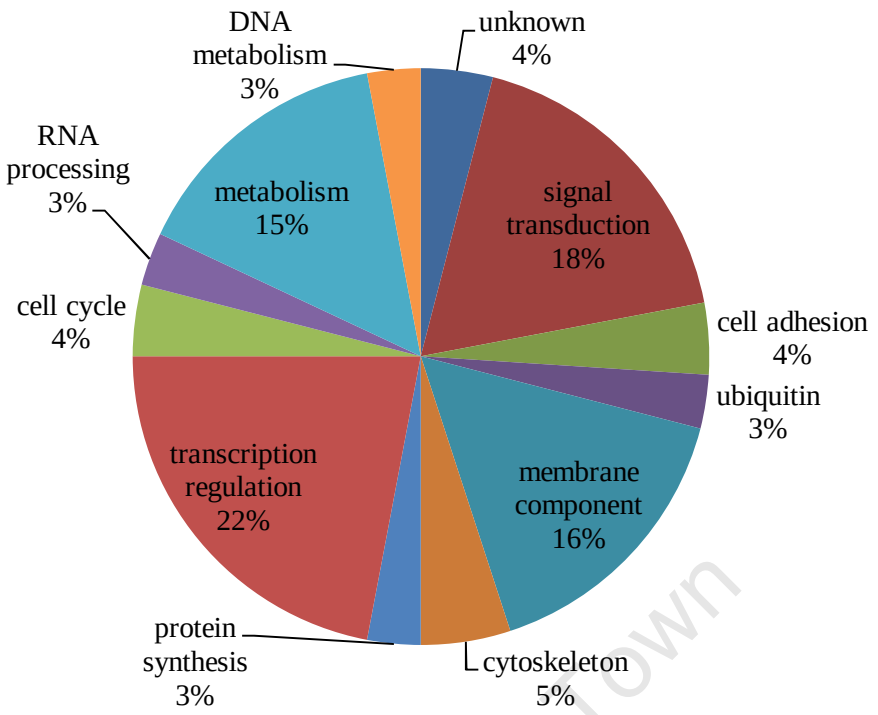


Figure 1.3 Pie chart illustrating the diverse protein functions of the 87 known XLMR genes. (Adapted from Chiurazzi et al. 2008).

1.3 THE EPIGENETIC BASIS OF DISEASE

Epigenetics is defined as mitotically and meiotically heritable changes in gene regulation which are not dependant on the primary DNA code (Levenson and Sweatt. 2005). The epigenetic mechanisms which govern gene regulation consist primarily of two components: DNA methylation and histone modifications, which act in a co-ordinated fashion with non-coding RNA (ncRNAs) and chromatin remodelling complexes. These mechanisms function together to create a dynamic chromatin structure which controls gene expression in a tissue specific and developmentally regulated manner. In addition, epigenetic responses can be articulated in the presence of both stochastic and environmental cues further lending to the flexibility of this regulatory mechanism.

The role of epigenetics in disease is a rapidly expanding concept with a number of diseases resulting from aberrant epigenetic mechanisms (Table 1.4). To date, the best characterised is the role of aberrant DNA methylation in the development of numerous cancers, as reviewed elsewhere by (Sharma, Kelly and Jones. 2010). However, aberrant epigenetic mechanisms are also involved in a variety of disorders of which MR is a feature (Table 1.4).

Table 1.4 Disorders associated with aberrant epigenetic mechanisms

Disorder	Manifestations	Etiology
Angelman syndrome	MR	Imprinting deregulation of *UBE3A* at 15q11-13 (maternal) (Bienvenu and Chelly. 2006)
Prader-Willi syndrome	MR	Imprinting deregulation of genes at 15q11-13 (paternal) (Bienvenu and Chelly. 2006)
ATR-X syndrome	MR	*ATRX* mutations →hypomethylation of specific repeat and satellite sequences (Gibbons et al. 2000)
Coffin-Lowry syndrome	MR	*RSK2* mutations → histone phosphorylation (Harum, Alemi and Johnston. 2001)
Rett syndrome	MR	*MECP2* mutations → recruits repressive and activating complexes to methylated DNA (Chahrour and Zoghbi. 2007, Yasui et al. 2007)
Rubinstein-Taybi syndrome	MR	*CBP* mutations → histone acetylation (Roelfsema et al. 2005)
ICF syndrome (immunodeficiency, centromeric region instability and facial anomalies syndrome)	Chromosome instability, MR	*DNMT3B* mutations →DNA hypomethylation (Ehrlich et al. 2008)
BWS (Beckwith-Wiedemann syndrome)	Organ overgrowth	Imprinting deregulation of genes at 11p15.5 (Maher and Reik. 2000)
Leukaemia	Disturbed haematopoiesis	Chromosomal translocations involving histones acetyl transferases (HATs) and histone methyltransferases (HMTs) (Claus and Lubbert. 2003)
Various cancers	Microsatellite instability	De novo methylation of *MLH1*(Kane et al. 1997)
	Uncontrolled proliferation	De novo methylation of various gene promoters (Jones and Baylin. 2002)
	Disruption of SWI-SNF chromatin remodelling complex	*SNF5*, *BRG1*, *BRM* mutations (Reisman, Glaros and Thompson. 2009)
	Overexpression of *IGF2*, *CDKN1C* silencing	Loss of imprinting (Chao and D'Amore. 2008, Soejima et al. 2004)

1.3.1 CHROMATIN STRUCTURE

The human genome comprises of approximately three billion base pairs. These nucleotides, which when extended in linear form would be 2m in length, need to be tightly compacted within the cell which has a diameter of only 10µm. This excessive DNA packaging into the cell is achieved through the formation of chromatin. By definition, chromatin is the state in which DNA is packaged within the cell and comprises DNA as well as associated proteins and RNA.

The most basic unit of chromatin is the nucleosome (Figure 1.4). Each nucleosome is linked to the next by a short fragment (10-100bp) of linker DNA completing a 10nm fibre representing a 'beads on a string' structure. Furthermore, the histone protein, H1, anchors the two turns of DNA around the nucleosome and adjacent H1 histones associate with one another to make up the 30nm solenoid structure. Additional hierarchical packaging of the DNA results ultimately in the chromosomal structure visible at metaphase (Figure 1.5) (Alberts. 2008).

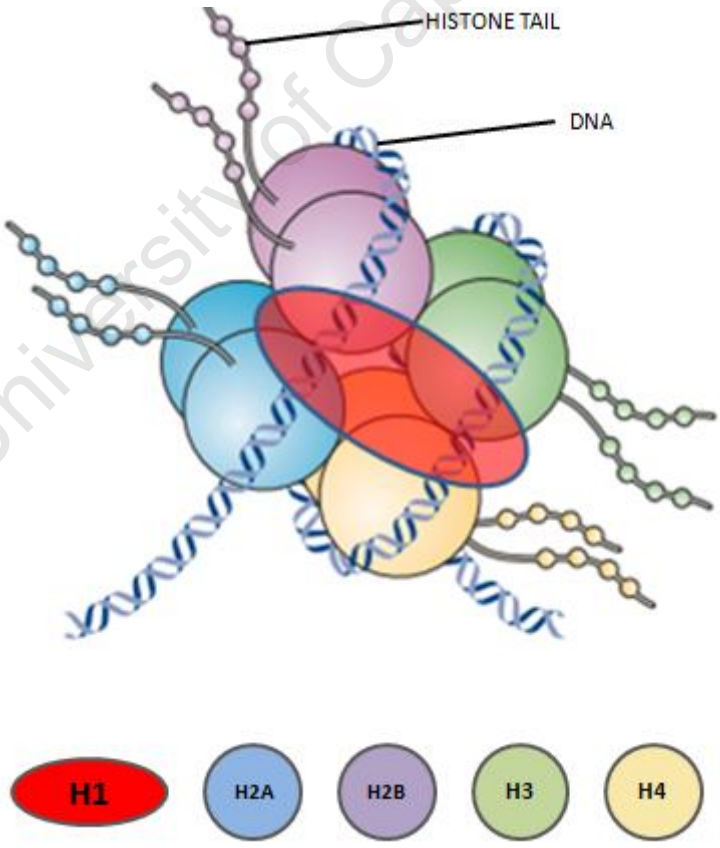


Figure 1.4 Structure of the nucleosome. The nucleosome consists of 147bp of DNA wrapped around the histone protein octomer which consists of two molecules of each of the four core histones (H2A, H2B, H3 and H4) is anchored by the histone H1. The amino acid residues within the histone tails are amenable to modifications which alter the tertiary chromatin structure upon which transcription is dependant (adapted from Levenson and Sweat, 2005).

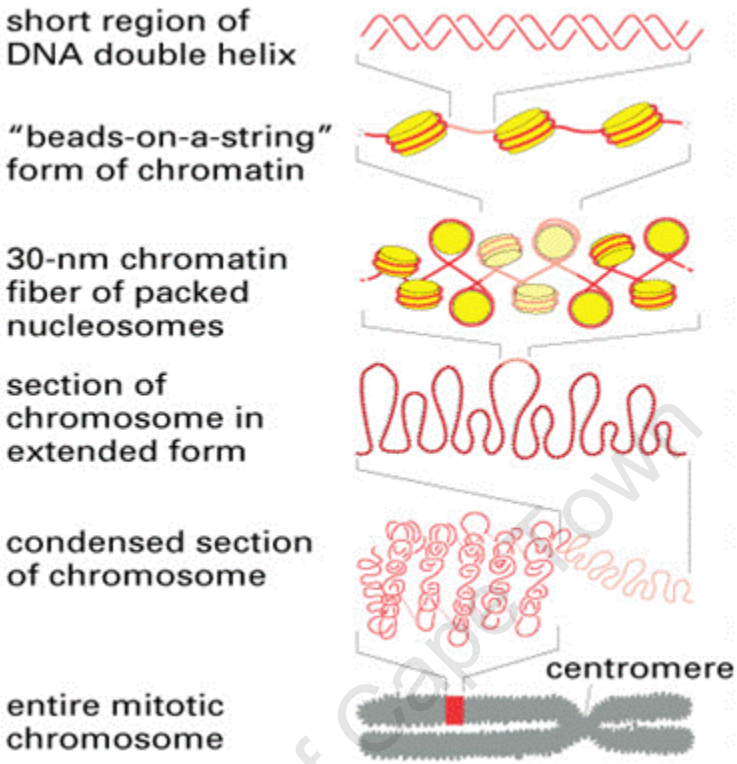


Figure 1.5 The hierarchical process of packaging DNA into the cell (Alberts et al. 2008).

Euchromatic regions are generally regions of DNA associated with an open chromatin formation and are actively transcribed. Conversely, heterochromatin comprises tightly packaged DNA in a 'closed' chromatin structure and encompasses regions that are not being transcribed. This dynamic nature of chromatin is achieved by epigenetic modifications which constitute the 'epigenome' and consists primarily of two mechanisms: histone modifications and DNA methylation.

1.3.2 THE EPIGENETIC MECHANSIMS

1.3.2.1 Histone modifications

The four core histones which make up the nucleosome are generally globular in shape but are also characterised by N-terminal 'tails'. The amino acid residues that constitute these histone tails protrude beyond the nucleosomal structure and are amenable to covalent post-translational modifications. These modifications alter the

generally basic nature of the nucleosome and hence affect the ability of the structure to bind negatively charged DNA thus inducing either a 'closed' or 'open' conformation. In addition, these diverse histone modifications act as epigenetic marks which signal downstream protein-protein interactions which further maintain the appropriate chromatin state.

The complexity of histone modifications is illustrated by the following features:

- The 60 different amino acid residues within the histone tails are each amenable to various covalent modifications. (Kouzarides. 2007) (Table 1.5). Given the octomeric composition of the nucleosome this equates to a total of at least 480 residues amenable to post-translational modifications per single nucleosome. The most commonly targeted residues are lysine and arginine, with less well-studied modifiable residues including serine, threonine, glutamic acid and proline.
- The array of histone modifications which can occur, of which histone acetylation, methylation and phosphorylation are the best categorised. Less well known modifications include ubiquitylation, sumoylation, ADP ribosylation, deimination and proline isomerisation.
- The variable nature of Lysine methylation which can occur in either mono-, di- or tri-methyl forms and arginine methylation can include mono- and di- (both asymmetric and symmetric) methylation.
- The recruitment of a wide-range of non-histone proteins by these histone modifications concomitantly signals for the ordered recruitment of additional gene-regulatory epigenetic complexes (as reviewed in Kouzarides. 2007). These enzyme complexes include the ATPase dependant remodelers, the silencing polycomb group proteins (PcG) and the activating trithorax group proteins (TtxG) (Schuettengruber et al. 2007).

Table 1.5 Some examples of well known histone modifications and their effect on the chromatin state (adapted from Kouzarides. 2007).

Chromatin modification	**Enzymes involved**	**Example of modifiable amino acid**	**Effect on transcription**
Acetylation	Histone acetyltransferases (HATs) Histone deacetyltransferases (HDACs)	H2A, H2B, H3, H4	Acetylation →activation deacetylation →repression
Methylation	Histone methyltransferases (Lysine residues) (HMTs) Histone demethyltranferases (e.g. *KDM5C*)	Methylation at H3K9 H3K27 H3K4 H3K36	 repression repression activation activation
Phosphorylation	e.g. RSK2	H3S10	activation
Ubiquitylation	e.g. Bmi/Ring1A protein found in the polycomb complex	H2AK119	repression

K= Lysine, S= Serine

1.3.2.2 DNA methylation

DNA methylation is the only known covalent modification of DNA in humans and is the result of the addition of a methyl group to the 5th position of the cytosine pyrimidine ring located at CpG dinucleotides (Wu and Santi. 1985). These CpG dinucleotides are underrepresented in the genome and methylation occurs at up to 70% of these sites (Cooper and Krawczak. 1989). The remainder are generally unmethylated and are organised in clusters known as CpG islands which encompass gene promoters (Bird. 1986).

DNA methylation is catalysed by enzymes known as DNA methyltransferases (DNMTs), details of which appear in Table 1.6. DNA methylation was long believed to be a stable epigenetic mark; however, the observation of demethylation during development suggests that a DNA demethylation enzyme/s may also exist. This is a subject of much debate with the alternative proposal that DNA demethylation may be a passive event occurring during DNA replication (Costello and Plass. 2001). However, rapid DNA demethylation has been observed in long term memory formation (Miller and Sweatt. 2007). Also more recently, a transcriptionally active gene (*TFF1*) that undergoes promoter DNA methylation and demethylation in an oestrogen-dependant manner has been described (Kangaspeska

et al. 2008, Metivier et al. 2008). Finally, the Vitamin D biosynthesis gene *CYP27B1* undergoes active DNA de/methylation in the presence of parathyroid hormone (PTH) and vitamin D3 respectively (Kim et al. 2009). Collectively, this evidence suggests that DNA demethylation may, in part, be responsible for temporal and spatial gene expression.

Table 1.6 Classes of human DNA methyltransferases and their function.

DNA methyltransferases	**Function**
DNMT1	Maintenance of the parental DNA methylation pattern during DNA replication(Razin and Szyf. 1984) Forms a repressive complexes including HDACs (Robertson et al. 2000)
DNMT3A	De novo DNA methylation capabilities Establish methylation patterns during development Able to target unmethylated CpG islands (Okano et al. 1999).
DNMT3B	De novo DNA methylation capabilities Establish methylation patterns during development Able to target unmethylated CpG islands (Okano et al. 1999).
DNMT3L	Lacks catalytic enzyme activity, but could act as a DNMT3A/B co-factor in the recruitment of HDAC (Deplus et al. 2002). Essential for genomic imprinting during germ cell development (Bourc'his et al. 2001).
DNMT2	Lacks catalytic activity, function unclear, able to methylate a tRNA (Goll et al. 2006, Okano, Xie and Li. 1998)

DNA methylation has a number of diverse functions:

- **Transcriptional silencing**: A well-established attribute of cytosine methylation is transcriptional silencing (Razin and Riggs. 1980), achieved by two mechanisms. Firstly, cytosine methylation can prevent the DNA binding of transcription factors to regulatory regions, hence inhibiting recruitment of transcription-inducing complexes (Watt and Molloy. 1988). Secondly, the recruitment of methylated DNA binding domain (MBD) proteins to methylated CpG islands leads to the recruitment of co-repressors and histone modifying enzymes which reinforce the inactive chromatin state at target genes (Boyes and Bird. 1991, Fujita et al. 1999, Hendrich and Bird. 1998, Jones et al. 1998)

- **X-inactivation**: in females the active X chromosome is relatively hypomethylated at promoter regions and hypermethylated at intragenic regions (Hellman and Chess. 2007).
- **Genomic imprinting**: variable DNA methylation of paternal and maternal promoter regions controls monoallelic expression at target loci.
- **Genome stability:** is enforced by the hypermethylated DNA state as illustrated by the pathogenesis of the ICF syndrome. This disorder results from *DNMT3B* mutations which predispose the resultant hypomethylated centromeric regions to chromosomal translocations (Ehrlich. 2003).
- **Suppression of transposable elements**: LINEs, SINEs and intracisternal A particle (IAP) elements have the potential to transpose throughout the genome and hence disrupt genomic integrity. These elements are subject to DNA methylation in order to prevent expression and subsequent mobilisation (Rollins et al. 2006).
- **Embryological development**: DNA methylation plays a critical role during this phase of human development, particularly during implantation when de novo demethylation is succeeded by establishment of DNA methylation patterning (Kafri et al. 1992, Monk, Boubelik and Lehnert. 1987).

1.3.2.3 Interplay between DNA methylation and histone modifications

DNA methylation and histone modifications both have regulatory roles in the mediation of the chromatin state. However, these mechanisms are not independent and correct maintenance of a particular expression state is dependent on a functional interaction/relationship between these two epigenetic mechanisms. There has been much debate about the nature and direction of this relationship. Given the evidence for both histone modifications and DNA methylation acting as the primary epigenetic mark, it is likely that a bi-directional model will be adopted.

There are instances where DNA methylation acts as the primary epigenetic mark. For instance, DNA methylation acts as a marker for the recruitment of MBD proteins. These MBD proteins recruit additional chromatin repressive complexes including HDACs and members of the polycomb repressive complexes. There are numerous examples of this association with the most well studied perhaps being the

transcription regulation complexes involving MeCP2. MeCP2 a MBD protein, has numerous regulatory functions which include transcription repression e.g. at the *BDNF* promoter (Martinowich et al. 2003). This repression acts by binding to methylated DNA at promoter regions and recruiting members of a repressive complex including SIN3A (transcriptional co-repressor), BRM and ATRX (SWI/SNF- related-chromatin-remodelling proteins) and HDACs (Bienvenu and Chelly. 2006, Nan et al. 2007).

Conversely, histone modifications may act as a marker for DNA methylation. This was elegantly illustrated by Mutskov et al. (2004) where transgenes integrated into chicken embryos were silenced by epigenetic mechanisms. This transgene silencing by promoter DNA methylation was preceded by histone acetylation and H3K4 demethylation, suggesting that, in this instance, DNA methylation is not the primary epigenetic mark (Mutskov and Felsenfeld. 2004). Also, a recent genome-wide promoter analysis of the epigenetic changes between embryonic stem cells (ESCs), neuronal progenitor cells and terminally differentiated pyramidal neurons found that genes which were destined to become inactive in the committed state were inactive prior to DNA methylation in 62% of cases. This suggests that the inactive state is conferred by an additional epigenetic mark, possibly H3K4 demethylation and H3K27 methylation (Mohn et al. 2008).

These examples illustrate just some of the incidences of a bidirectional relationship between DNA methylation and histone modifications and further reinforce the notion that there is an epigenetic interplay between these two mechanisms.

1.3.3 EPIGENETICS IN NEUROGENESIS, NEURONAL DEVELOPMENT AND FUNCTIONING

There is ample evidence for the role of epigenetic mechanisms controlling both neurogenesis as well as neuronal development and functioning. Neurogenesis is, at least in part, controlled by the protein, restriction element 1 (RE1) silencing transcription factor (REST). In non-neuronal tissues, REST binds to DNA at short consensus sequences known as RE1 (or neuron-restrictive silencing element (NRSE) and promotes protein-protein interaction with DNMTs, HDACs, MBDs, transcription factors, chromatin remodelers and co-regulators (coREST). The formation of this REST-associated protein complex enforces a repressive epigenetic state which includes marks such as: DNA methylation, H3K9 methylation and H3K4 demethylation at the target DNA region (Ballas and Mandel. 2005). In this way neuronal gene expression is restricted to the nervous system. Conversely, in stem cells as well as neuronal progenitor cells, the REST complex bound at the RE1 motif of neuronal promoters is generally associated with active chromatin modification marks. Rapid dissociation of REST from the RE1 site during neurogenesis is achieved through REST degradation or ncRNA-mediated activation, hence allowing neuronal gene expression (Ballas and Mandel. 2005, Kuwabara et al. 2004).

It has also been hypothesised that epigenetic mechanisms play a role in learning and memory. In order to develop long-term memories there is accumulating evidence that neuronal functioning makes use of epigenetic mechanisms in a manner similar to cell state commitment during development (Figure 1.6) (Levenson and Sweatt. 2005, Levenson et al. 2006, Miller, Campbell and Sweatt. 2008, Swank and Sweatt. 2001). Again, these epigenetic changes include both changes in DNA methylation and histone modifications as well as interaction between the two. For example, it was demonstrated that acetylation of histone H3 occurs in rat hippocampal regions subsequent to contextual fear conditional training (Levenson et al. 2004). In addition, DNMT inhibition prevented this memory-associated H3 acetylation, thus illustrating the necessity for interplay between these two epigenetic mechanisms for memory formation (Miller, Campbell and Sweatt. 2008).

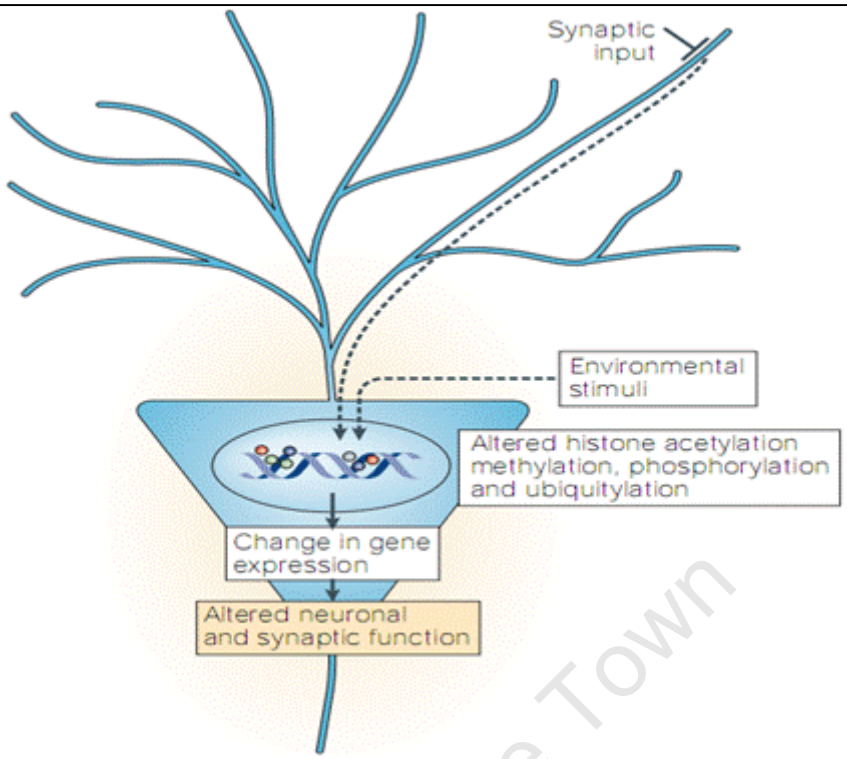


Figure 1.6 Epigenetic mechanisms in learning and memory. In response to a particular external stimulus, epigenetic changes such as DNA methylation and histone modifications lead to a change in gene expression. This modified expression ultimately leads to altered neuronal and synaptic function which forms the basis of learning as well as memory formation and consolidation (adapted from Levenson and Sweat, 2005)

1.3.1 PATHOGENEIC ROLE OF EPIGENETICS IN XLMR

Perhaps the best example of the role of epigenetics in neuronal development and functioning is the observation of cognitive defects in patients with mutations in XLMR genes that play a role in transcriptional regulation. Of the 87 known XLMR genes, 22% are either known or hypothesised to play a role in transcriptional regulation (Table 1.7).

Table 1.7 XLMR genes involved in transcription regulation with putative/proven epigenetic effects (adapted from Chiurazzi et al. 2008)

	Gene	Function
1.	*ARX*	Transcription repressor and activator (Seufert, Prescott and El-Hodiri. 2005)
2.	*ATRX*	Chromatin remodeler
3.	*BCOR*	Chromatin remodeler
4.	*BRWD3*	Chromatin remodeler (Field et al. 2007)
5.	*CDKL5*	Kinase, involved in MeCP2 phosphorylation
6.	*JARID1C*	Histone H3K4 demethylation (Iwase et al. 2007)
7.	*MECP2*	MBD protein, transcription repression and activation (Chahrour and Zoghbi. 2007)
8.	*MED12*	RNA Polymerase II transcription mediator, transcription co-activator
9.	*PHF6*	Transcription regulation
10.	*PHF8*	Transcription regulation
11.	*PQBP1*	Transcription regulation
12.	*RPS6KA3*	Kinase
13.	*RPS6KA6*	Kinase
14.	*SIZN1*	Transcription regulation (Cho et al. 2008)
15.	*SOX3*	Transcription regulation
16.	*ZNF41*	Transcription regulation: DNA binding protein
17.	*ZNF674*	Transcription regulation: DNA binding protein
18.	*ZNF711*	Transcription regulation: DNA binding protein
19.	*ZNF81*	Transcription regulation: DNA binding protein

The most well-studied of these 'epigenetic' XLMR genes is *MECP2*, mutated in patients with NS-XLMR, Rett syndrome (atypical and typical) and an Angelman syndrome-like phenotype (Couvert et al. 2001, Villard et al. 2000, Watson et al. 2001). MeCP2 is a MBD protein originally thought to act as a transcriptional repressor by binding methylated DNA and recruiting chromatin repression complexes (Chahrour and Zoghbi. 2007). While this transcription repressor activity of MeCP2 has been reported e.g. at the *BDNF* gene (Chen et al. 2003), MeCP2 also acts as a transcription activator. In mouse hippocampal cells MeCP2 acts as an activator at 85% of investigated gene promoters by association with the protein cAMP responsive element binding protein 1 (CREB1) at the promoter start site of active genes (Chahrour et al. 2008). Finally, patients with *MECP2* mutations have aberrant dendritic branching and an altered number of synapses suggesting defective

postnatal neuronal function (LaSalle et al. 2001). Collectively, these results suggest that aberrant expression of MeCP2 has effects on expression of downstream target genes ultimately leading to the neurological defects in patients with MeCP2 mutations.

While less well-studied, evidence suggests that the additional XLMR 'epigenetic' genes may lead to neurological defects due to aberrant gene expression of target genes and that this aberrant expression is a downstream effect of alteration to epigenetic mechanisms. The focus of this study was therefore to investigate DNA methylation profiles in XLMR 'epigenetic' gene mutation positive individuals. The outcomes of these experiments contribute to the understanding of the role of DNA methylation in XLMR pathogenesis. However, in addition, the development of this whole-genome DNA methylation approach presents an alternative strategy for the identification of novel MR candidate genes.

1.4 AIMS AND OBJECTIVES

Aim: To establish whether an alteration to the DNA methylation profile is a molecular feature of MR in patients positive for mutations in putative 'epigenetic' XLMR genes.

This aim was achieved by fulfilment of the following **objectives** (Figure 1.7):

Part A: Identifying the primary DNA mutation by:

- Whole-genome CNV analysis, which included:
    - CNV analysis in a clinically stratified cohort using whole-genome SNP arrays
    - Putative CNV validation from the SNP array using genomic qPCR (quantitative PCR)
    - Assessing the association between the detected CNV and MR, using a variety of techniques as described above

- Linkage analysis and positional candidate gene screening, which involved:
    - Whole X chromosome microsatellite genotyping for linkage analysis in suitable XLMR families
    - Positional candidate gene screening within the linked critical intervals to identify DNA sequence alterations using Sanger DNA sequencing
    - Assessing the association between detected DNA sequence alterations and XLMR using a variety of techniques including: bioinformatic interrogation, literature review, mRNA investigation, familial co-segregation and background population screening.
- Functional candidate genes analysis which included:
    - Screening of functional candidate genes for DNA sequence alterations in clinically stratified cohorts using Denaturing high performance liquid chromatography (dHPLC) and Sanger DNA sequencing,
    - Assessment of the association between detected DNA sequence alterations and XLMR using a variety of techniques as described above.

Part B: Identifying whole-genome DNA methylation aberrations by:

- Investigating the genome-wide DNA methylation patterns in XLMR gene mutation positive individuals using MeDIP (methylation dependant immuno-precipitation) –Chip
- Validating putative differentially methylated regions (DMRs) by bisulphite Sanger DNA sequencing
- Semi-quantification of DNA methylation ratios at DMRs using PCR cloning of bisulphite treated products and subsequent DNA sequencing of clones.

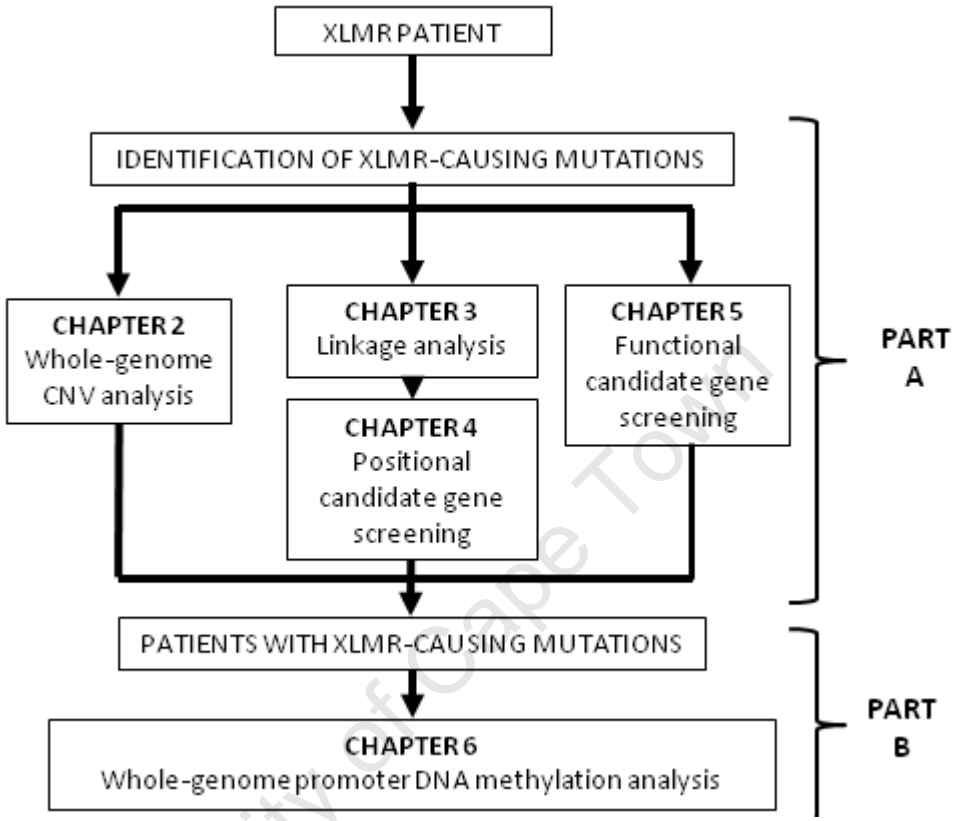


Figure 1.7 The strategic approach adopted for fulfilment of the objectives of this study

PART A

INTRODUCTION

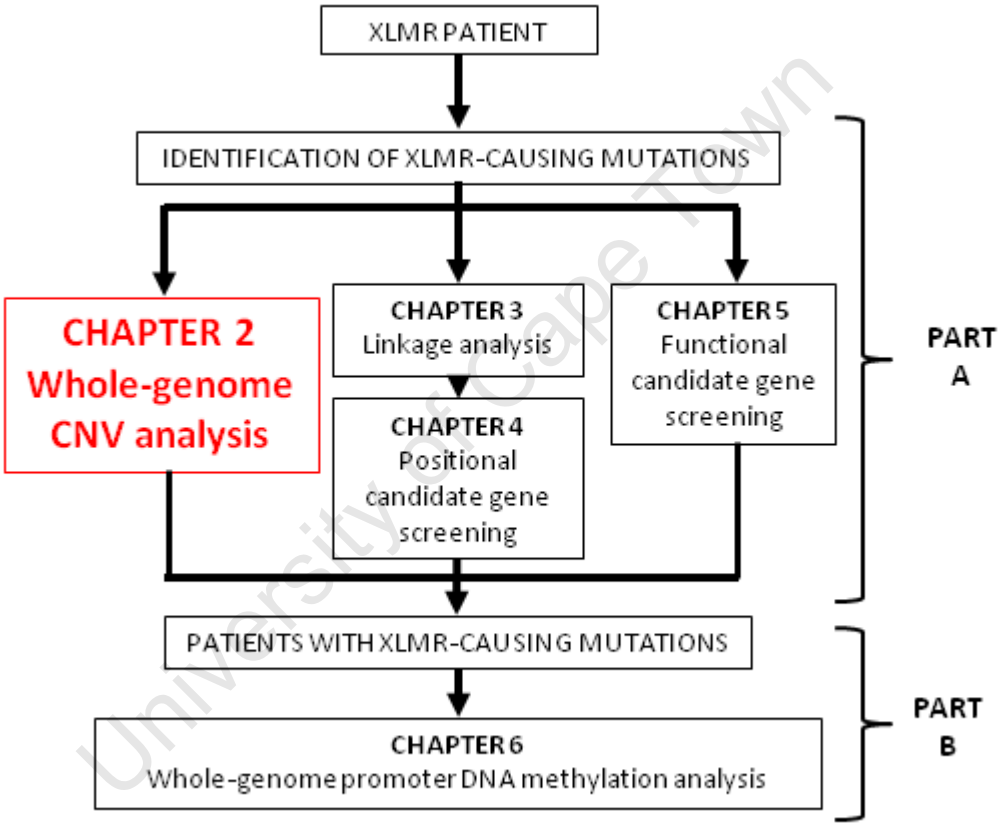
IDENTIFICATION OF DISEASE-CAUSING MUTATIONS IN SOUTH AFRICAN XLMR PATIENTS

In order to investigate the central hypothesis of this dissertation, that alteration to the DNA methylation profile is a molecular feature of MR in patients positive for mutations in putative 'epigenetic' XLMR genes, it was first necessary to identify patients with mutations in these XLMR genes. Given the complex genetic landscape of XLMR disease-causing alterations, a multifaceted approach was required in order to define the molecular etiology of XLMR. This approach was a vital first step in identifying suitable XLMR mutation positive patients in whom this hypothesis could be tested.

Part A of this dissertation details the various molecular techniques and strategic approaches employed in order to ascertain the underlying molecular etiology in this cohort of South African XLMR patients. These techniques include copy number analysis using SNP arrays to identify disease-causing CNVs across the genome (Chapter 2). Also, X-chromosomal linkage analysis (Chapter 3) and subsequent positional candidate gene screening (Chapter 4) was conducted. Finally, functional candidate XLMR genes were selected for investigation in stratified cohorts based on mutation frequencies previously reported by the international community (Chapter 5). Those patients shown to be positive for mutations in XLMR genes were subsequently prioiritised for investigation of whole-genome DNA methylation pattern aberrations (Part B).

CHAPTER 2

WHOLE-GENOME COPY NUMBER VARIATION (CNV) ANALYSIS



2.1 INTRODUCTION

Alterations of chromosomal copy number have long-been established as a major contributor to MR. Traditionally, cytogenetic techniques such as G-banded karyotyping have been used to identify large chromosomal abnormalities in 3-5% of MR patients (Stankiewicz and Beaudet. 2007). However, due to the limited resolution of this technique (5-10Mb), a substantial proportion of aberrations go undetected. Higher-resolution technologies were thus developed to address this limitation. These technologies included molecular cytogenetic techniques such as FISH and molecular techniques including MLPA. These target-specific technologies have been successfully utilised for the detection of common microdeletion/duplication syndromes and subtelomeric rearrangements in 5-6% of MR cases (Flint and Knight. 2003). However, to date, the most promising technological advancement has been the adaption of microarrays to detect CNVs at an even greater resolution, with detection rates of up to 24% in MR patients (Aston et al. 2008).

2.1.1 MICROARRAY TECHNOLOGIES

Early application of microarrays for CNV detection using bacterial artificial chromosome (BAC) arrays was to assess copy number in solid tumour cancer cells (Kononen et al. 1998). Subsequently, this technology has been applied in a wider diagnostic setting, particularly in the molecular diagnosis of MR. Recent technical developments within this field have led to the replacement of BAC probes with synthesised oligonucleotides. These oligonucleotide arrays have a higher resolution, greater flexibility and are less time-consuming to manufacture. There are two primary applications of oligonucleotide microarrays for CNV detection which differ primarily on the reference genome source used to calculate copy number:

- **Two-fluorophore hybridisation**: entails competitive hybridisation of differentially fluorescently-labelled test and reference samples to the array platform (generally arrays comprised of random 50-80mer probes specific to the regions of interest are used). Subsequently, allelic copy number is calculated from the fluorescent signal intensity ratios at each locus.

- **One- fluorophore hybridisation**: a single fluorescently-labelled test sample is hybridised to the array platform (generally arrays comprised of SNP-specific probes are used). Subsequently, the intensity ratios between a test and reference sample are compared in silico to calculate allelic copy number.

The second method has the advantage of producing both genotype and copy number data, also, it negates the need for the use of reference DNA for every sample tested (for a more detailed review see Friedman et al. 2006, Zahir and Friedman. 2007).

2.1.2 CNV DETECTION RATES IN MR STUDIES USING MICROARRAYS

CNV detection using both two-fluorophore, and one-fluorophore microarray hybridisation yield variable detection rates of 5-24% (Table 2.1). These pick-up rates were dependent on both the resolution of the array platform as well as the clinical inclusion criteria (Table 2.1). Analyses of these various studies indicate that the highest diagnostic yields were obtained in those studies which employed high-resolution whole-genome microarrays. Also, higher pick-up rates were demonstrated in cohorts consisting of patients presenting with MR in conjunction with dysmorphism and/or additional features, or MCA.

Detection rates were furthermore largely dependent on the microarray design, more specifically, whether a 'tiling' or 'targeted' array was used. There has been much debate in recent literature regarding the use of 'targeted' versus 'tiling' microarrays to detect disease-causing CNVs in MR cohorts (Bejjani and Shaffer. 2006, Veltman and de Vries. 2006, Veltman and de Vries. 2007). 'Targeted' arrays are designed to detect disease-causing aberrations in subtelomeric/ pericentromeric regions as well as chromosomal intervals commonly implicated in microdeletion/duplication syndromes. The major advantage of these arrays is the exclusion of detecting variants with unknown significance as interpretation of the disease-association of some CNVs can be difficult.

Table 2.1 Details of previous studies that employed microarrays to detect genomic copy number changes in MR patient cohorts.

Study	Array platform used	Clinical inclusion criteria	# patients tested	# disease-causing mutations	Pick-up rate (%)
(Vissers et al. 2003)	3569 BAC clone whole-genome tiling array	MR with dysmorphism	20	2	10
(Shaw-Smith et al. 2004)	3500 clone BAC whole-genome tiling array	MR with dysmorphism and/or additional features	50	7	14
(de Vries et al. 2005)	32447 BAC clone whole-genome tiling array	MR	100	10	10
(Schoumans et al. 2005)	2600 BAC clone whole-genome tiling array	MR with dysmorphism	41	4	10
(Tyson et al. 2005)	Spectral genomics 3Mb and 1Mb BAC arrays	MR with dysmorphism	22	3	14
(Ballif et al. 2006)	831 or 969 BAC clone targeted array	Diverse DD- MR	3600	182	5
(Friedman et al. 2006)	Affymetrix 100K oligonucleotide array	MR	100	11	11
(Krepischi-Santos et al. 2006)	3500 clone BAC whole-genome tiling array	Syndromic MR or additional features	95	16	17
(Menten et al. 2006)	3431 clone BAC whole-genome tiling array	MCA and/or MR	140	19	14
(Miyake et al. 2006)	2173 clone BAC whole-genome tiling array	Idiopathic MR	30	5	17
(Rosenberg et al. 2006)	3500 clone BAC whole-genome tiling array	MR with dysmorphism	81	13	16
(Shaffer et al. 2006)	831 BAC clone targeted array	Diverse DD- MR	1500	84	6
(Sharp et al. 2006)	2007 BAC clone array encompassing 130 LCR regions	MR and/or dysmorphism or MCA	290	16	6
(Engels et al. 2007)	6K BAC array and 8K BAC array	MR with dysmorphism and/or additional features or MCA	16 (6K array) 44(8K array)	1 5	6 11
(Fan et al. 2007)	434 BAC clone targeted array (Constitutional Chip – Spectral genomics) Whole-genome oligonucleotide array (Agilent) with 30-35kb resolution	MR-DD with two or more dysmorphic features	100 (targeted array) 100 (whole-genome)	5 15	5 15
(Hoyer et al. 2007)	100K Affymetrix oligonucleotide array	MR with/out MCA	104	10	9

Study	Array platform used	Clinical inclusion criteria	# patients tested	# disease-causing mutations	Pick-up rate (%)
(Lu et al. 2007)	366 or 853 BAC clone targeted array	Diverse DD- MR	749 (366-BAC array) 1695 (853-BAC array)	46 125	6 7
(Shaffer et al. 2007)	589 BAC clone targeted array	Idiopathic MR	8789	604	7
(Shen et al. 2007)	>10 000 60-mer Agilent oligonucleotide targeted array (to known MR regions)	DD-MD with dysmorphism or MCA	211	25	12
(Wagenstaller et al. 2007)	100K Affymetrix oligonucleotide array	MR with/out dysmorphism	67	11	16
(Aston et al. 2008)	Spectral genomics/ PerkinElmer oligonucleotide Constitutional Chip (targeted array) Spectral Chip 2600 with a 1Mb resolution (whole genome array)	MR	309 (targeted array) 245 (whole genome) 522 (both arrays)	27 58 97	9 24 19
(Koolen et al. 2008)	32447 BAC clone whole-genome tiling array	MR	386	35	9
(Nowakowska et al. 2008)	853 BAC clone targeted array	MR with additional features	91	19	12
(Friedman et al. 2009)	500K Affymetrix oligonucleotide array	MR with dysmorphism	100	16	16
(Jaillard et al. 2009)	Agilent 44K oligonucleotide arrays	MCA and/or MR	132	19	14
(McMullan et al. 2009)	500K Affymetrix oligonucleotide array	Unexplained MR	120	18	15

DD- developmental delay, MCA – multiple congenital anomalies

In contrast to 'targeted' arrays, whole genome 'tiling' arrays present an unbiased approach to CNV detection, with both benign and disease-causing variants being identified. In general, explication of the pathogenic nature of a CNV is twofold. Firstly, the presence of the CNV in databases such as the database of genomic variation (DGV) or in the literature (also catalogued in the UCSC Genome Browser as 'structural variation tracks') would suggest the CNV is benign. Secondly, parental DNA samples are assessed to determine whether the CNV arose de novo or was inherited from an unaffected parent. However, this exclusion rationale also has its caveats in that it assumes that a particular CNV is fully penetrant. Also, the extent of CNV polymorphism is at present, unknown, concomitantly a number of benign variants have yet to be characterised and catalogued in the aforementioned databases.

Despite the caveats associated with CNV interpretation, a whole-genome 'tiling' approach has been shown to have a higher diagnostic yield as compared to 'targeted' arrays (Table 2.1). In a recent study of over 1000 MR patients using array-technologies, Aston and colleagues (2008) demonstrated that the detection rate of 'targeted' and 'tiling' arrays was 9% and 24% respectively (Aston et al. 2008). Similarly, a comparative study of 100 MR patients identified disease-causing variants in 5% of patients using a targeted array, compared to a 15% pick-up rate using whole-genome tiling arrays (Fan et al. 2007). The higher detection rate in whole-genome studies can be ascribed to two main attributes; firstly a number of disease-causing CNVs arise de novo. These de novo mutations are generally not located in CNV hotspots and are thus not due to non-allelic homologous recombintation (NAHR) which tends to occur at low copy repeat (LCR) regions. Therefore, while these events tend to be rare, the occurrence of de novo disease-causing CNVs are relatively common, given the large size of the human genome. Secondly, the number of known microdeletion/duplication syndromes is, at present, under-represented. This is evident in the number of microdeletion/duplication syndromes that have been identified in recent years as a result of CNV analyses using microarrays. These novel microdeletion/ duplications include: 17q21.31 (OMIM **#610443**) (Koolen et al. 2008), 15q24 (Sharp et al. 2007), 14q11.2 (Zahir et al. 2007) 1q21.1 (OMIM **#612474**) (Mefford et al. 2008)

Given the high diagnostic yield of disease-causing CNVs in MR cohorts using microarray technologies, South African XLMR patients were selected from the Divisional DNA bank for copy number investigations. This genome-wide CNV analysis had a dual objective. Firstly, to assess the feasibility of incorporation of microarray CNV technologies into the diagnostic testing protocol offered to South African MR patients and their families. Secondly, in view of the central aim of this dissertation, to investigate DNA methylation profiles in XLMR mutation positive individuals, the presence of MR-causing CNVs in this cohort was investigated in an effort to identify individuals suitable for further epigenomic studies.

2.2 METHODS

Patient DNA samples selected for molecular investigations in this study, form part of the Division of Human Genetics, University of Cape Town (UCT) DNA bank. These probands were all referred by various clinicians to the Division for molecular testing for the fragile X syndrome (FXS). At the commencement of this dissertation DNA samples from approximately 450 families with MR had been catalogued and stored. In addition, patient/guardian permission was obtained for both the diagnostic testing of FXS as well as for further research (Appendix 2A) (subject to approval by the UCT Faculty of Health Sciences Ethics committee). All research was also approved by the aforementioned Ethics committee (REC 466/2005)

From the total of 450 MR families within the UCT DNA bank, 148 demonstrated an X-Linked inheritance pattern. Of these XLMR families, 35 (23%) tested positive for the CGG repeat expansion at the 5' UTR of the *FMR1* gene. Therefore, the molecular etiology of the remaining 77% of XLMR families or 90% of all patients in the DNA bank remained undefined.

2.2.1 PATIENT COHORT SELECTION

In total, 30 patients were selected for whole genome CNV investigations (Appendix 2B) based on the following inclusion criteria:

- MR and at least one additional clinical feature, dysmorphism or MCA.
- Negative for the *FMR1* expansion mutation
- Normal karyotype using standard G-banded cytogenetic analysis.

In this section of the dissertation, the study cohort comprised of both XLMR probands as well as isolated cases of MR due to the whole-genome nature of the CNV analysis (at the time of this study's commencement single chromosome (e.g. X-chromosome) oligonucleotide arrays were not yet available).

2.2.2 WHOLE-GENOME CNV ANALYSIS

A medium resolution (250K) whole genome array was utilised for CNV detection in this patient cohort. Given the lack of resources available at UCT, these studies were conducted together with Drs Hans van Bokhoven, Arjan de Brouwer and colleagues in the Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands.

2.2.2.1 Array preparation, labelling and hybridisation

Whole-genomic DNA was isolated from leukocytes of all 30 patients and subsequently analysed using the 250K Nsp Array [Affymetrix] as per manufacturer's instructions outlined in the GeneChip® Mapping 500K Assay Manual [Affymetrix]. Briefly, whole-genomic DNA from each patient was fragmented using the restriction enzyme (RE) NspI. Subsequently, the fragmented DNA was subject to whole-genome amplification using adapters ligated to the NspI cleavage sites. All samples were denatured, fluorescently end-labelled and hybridised to the array platform by incubation in a GeneChip® Hybridisation Oven 640 [Affymetrix]. Thereafter, all arrays were washed and stained in a GeneChip® Fluidics Station 450 [Affymetrix] and scanned with a GeneChip® Scanner 3000 7G [Affymetrix].

2.2.2.2 Array data analysis

The GeneChip® Operating Software (GCOS) [Affymetrix] was used to generate and acquire the raw data from each array. Raw data was analysed using the GeneChip® Genotyping Analysis Software (GTYPE) [Affymetrix] and genotype calls generated. In turn, these genotyping calls were used to determine copy number across the genome using the programme copy number analysis for genechip (CNAG) Version 2.0 as per software instructions (CNAG User Manual Version 2.0) (Yamamoto et al. 2007). A pool of ten female and ten male reference samples were used to determine copy number. From this pool, the CNAG software automatically selected the sex-matched reference (R) sample/s with the lowest standard deviation (optimally <0.2) as compared to the test (patient) (T) sample. The fluorescent signal intensity values from the selected reference sample/s were then used to calculate the Log2 transformed test over reference (T/R) value. These T/R values were visualised using the 'chromosome view' in CNAG as well as in chromosome-specific Microsoft Office Excel V.2007 [Microsoft] files so as to identify regions beyond the set thresholds of >0.3 and <-0.3 for duplications and deletions respectively.

2.2.2.3 Interpretation of CNVs

In order to limit the follow-up molecular investigations to those CNVs which were most likely associated with MR, variants were excluded according the criteria detailed below.

- **Previously identified in normal controls**: the CNV was excluded if it was previously reported in either the DGV at http://projects.tcag.ca/variation/ or in the 'structural variation tracks' at the UCSC Genome Browser http://genome.ucsc.edu/ (these tracks are compiled from the literature and encompass a variety of whole-genome CNV population-based studies)
- **Inheritance pattern**: if the family history was indicative of a particular MR inheritance pattern that the CNV did not adhere to, that CNV was excluded from further analysis (e.g. an autosomal CNV in a XLMR family)
- **No known genes**: If the CNV was predicted to encompass no known genes, it was excluded from further analysis, at this stage

2.2.3. CNV VALIDATION BY QUANTITATIVE REAL-TIME PCR

Subsequent to candidate CNV prioritisation, validation of all putative disease-causing CNVs was conducted using genomic quantitative real-time PCR (qPCR). For each putative CNV, five sets of primers were designed, three of which were located within the chromosomal region, while the remaining two flanked the region of interest (Appendix 2C). Furthermore, in patients Fx444.1 and XMR8.1 the breakpoints of the CNVs, Xq26.3-27.3Del and Xq25Dup respectively, were also determined using qPCR (Appendix 2C)

A qPCR mix was prepared consisting of: 8µM of forward and reverse primer, 5ng of genomic DNA, 1× iQ SYBRGreen Supermix [Biorad] and distilled water to a final volume of 25µl. The qPCR profile consisted of an initial denaturation step at 95°C for 3 minutes (min), followed by 40 amplification cycles at 95°C for 15 seonds (s) and 60°C for 30s. This was followed by incubation at 95°C for 1min and 65°C for an additional minute. Finally 60, ten second 0.5°C increment steps were performed (from a start point of 65°C) in order to create a dissociation (or 'melt') curve. All qPCR reactions were performed on a 7500 Fast Real-Time PCR system [Applied Biosystems]. DNA samples were amplified in triplicate for each primer pair of the qPCR reaction.

The $2^{-\Delta\Delta Ct}$ method was used to calculate fold differences (between calibrator and test sample) as described previously (Livak and Schmittgen. 2001). Amplicons from three reference genes, *CFTR*, *MTC8* and *STXBP5*, were used for quantification normalisation. Validation experiments demonstrated that the qPCR efficiency of amplicons from the reference genes and the target loci (chromosomal regions encompassing putative CNVs) was 100%, i.e. the amount of PCR product is doubled in every cycle. DNA from a normal female was used as the calibrator sample and the Cq (quantification cycle, or crossing point) value was adjusted for male samples when performing qPCR on X-chromosomal regions. The fold differences derived from the $2^{-\Delta\Delta Ct}$ method for each of the three reference genes (*CFTR*, *MTC8* and *STXBP5*) was averaged for each target primer pair and plotted on a line graph using Microsoft Office Excel V. 2007 [(Microsoft]. Also, where possible, the presence of the identified CNV was assessed in the proband's family members.

2.2.4 MOLECULAR FOLLOW-UP INVESTIGATIONS IN FAMILY FX444

2.2.4.1 X chromosome microsatellite marker genotyping in the family trio

In order to show X-chromosomal haplotype segregation in family Fx444, thirteen microsatellite markers distributed between 5-10Mb across the entire X chromosome were genotyped in all three family members (Fx444.1/2/3).

The microsatellite marker PCR amplification consisted of two rounds of PCR, firstly the microsatellite repeats were amplified with M13-labelled target-specific primers. Secondly, the PCR products were pooled and labelled with fluorescent tags (Hex, Ned, Fam, Rox) using primers specific for the M13 sequence. Also, the reverse primer contained a 'PIG' tail, used to reduce the stutter associated with microsatellite marker PCR products (Brownstein, Carpten and Smith. 1996). The PCRs were performed in a total volume of 25µl which consisted of 1× Taq buffer [Invitrogen], 75mM $MgCl_2$ [Invitrogen], 3µM dNTPs [Bioline] and 1.5U Taq [Invitrogen]. The two rounds of PCR differed in terms of both the template and primer used:

- PCR 1: 1µM forward/reverse M13-labelled microsatellite marker specific primers and 100ng of patient DNA (Appendix 2D)
- PCR2: 2µl of first-round PCR product was used as the template for amplification using 2µM fluorescently labelled M13 specific forward primer and 1µM PIG-tailed M13 specific reverse primer.

A touchdown PCR profile was used which consisted of: an initial denaturation at 95°C for 5min followed by 10 cycles of 95°C for 15s, primer annealing using a 'touchdown' protocol (0.5°C decrements per cycle) from 58°C to 54°C for 15s, followed by elongation at 72°C for 30s. Hereafter, 20/10 (PCR1/2) cycles of 95°C for 15s, 54°C for 15s, followed by elongation at 72°C for 30s and a final elongation step of 72°C for 10min were performed. PCR reactions were performed on an Applied Biosystems GeneAmp® PCR system.

Microsatellite marker PCR products were combined with the GeneScan™ 500 LIZ® size standard [Applied Biosystems] and subject to capillary electrophoresis on an ABI Prism® 3100 Genetic Analyser [Applied Biosystems] in accordance with manufacturer's protocols. Allele sizes were determined using the GeneMapper™

V3.0 [Applied Biosystems] software and haplotypes generated with the Cyrillic Software V2.0 [Cherwell Scientific].

2.2.4.2 X-inactivation investigations in unaffected carrier mother (Fx444.2)

The X-inactivation ratio in an unaffected carrier female (Fx444.2) was established using an X-inactivation assay. In order to conduct this assay, the methylation status of a CpG island located within the variable number of tandem repeats (VNTR) region of the *MAOA* gene was analysed as described previously (with modifications) (Hendriks et al. 1992, Plenge et al. 1999)

Methylation sensitive DNA digestion with HhaI

The cleavage activity of HhaI is blocked by methylation of cytosine nucleotides at CpG dinucleotides, which is a hallmark of X-inactivation. Therefore, DNA (1µg) from the unaffected mother (Fx444.2) as well as the sons (Fx444.1/3) was digested with the methylation sensitive enzyme, HhaI, as per manufacturer's instructions [Fermentas]. Simultaneously all DNA samples from family Fx444 were 'mock-digested' i.e. digestion reagents and incubation times were identical except for the lack of enzyme addition to the 'mock-digestion' samples.

PCR of MAOA VNTR region

The VNTR region of *MAOA* was subsequently amplified by PCR in all familial DNA samples (digested and mock-digested) (the heterozygosity of individual Fx444.2 at this locus was tested prior to HhaI digestion). The PCR was performed in a 25µl reaction consisting of: 10µM of forward (5' FAM-acagcctgaccgtggagaag 3') and reverse (5' gaacggacgctccattcgga 3') primers, 5µM dNTPs [Bioline], 100ng of DNA, 1× Go Taq buffer [Promega] and 1U Go Taq DNA polymerase [Promega]. The PCR profile consisted of an initial denaturation step at 95ºC for 5 min, followed by 30 cycles at 95ºC for 30s, 58ºC for 30s, and 72ºC for 30s. A final elongation at 72ºC for 5 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation].

Genotyping of the MAOA VNTR locus

The size of the *MAOA* VNTR PCR products was assessed by capillary electrophoresis conducted on an ABI Prism® 3100 Genetic Analyser [Applied Biosystems] in accordance with manufacturer's protocols. Software analysis was conducted using GeneScan™ V3.7 [Applied Biosystems]. Peak heights of fluorescence intensity from the GeneScan™ V3.7 software output were used to calculate allele-inactivation ratios.

Calculation of X-chromosome inactivation ratios.

In order to ensure success of the HhaI methylation-sensitive digestion, complete degradation of the male *MAOA* VNTR PCR product was first confirmed (this active X-chromosome is unmethylated and therefore not protected from HhaI digestion). Subsequently, the equation below (Equation 1) was used to calculate a corrected ratio (CR) for the two female alleles (adapted from Boudewijns, van Dongen and Langerak. 2007). The smaller of the two *MAOA* VNTR alleles was designated with the letter S and the larger allele, L. Bias towards either of the two alleles owing to PCR amplification bias was negated by division of the HhaI digested sample ratio (S1/L1) by the ratio of the mock-digested sample (S2/L2). From this CR, the percentage of inactivated X-chromosomes in the sample cell population for individual Fx444.1 was calculated.

Equation 1:

$$CR = \frac{S1\ (Hha1\ digested)}{L1\ (Hha1\ digested)} \div \frac{S2\ (Hha1\ un\text{-}digested)}{L2\ (Hha1\ un\text{-}digested)}$$

2.3 RESULTS

2.3.1 CNV DETECTION USING 250K SNP ARRAY

Whole-genome copy-number analysis of 30 MR patients with a syndromic clinical presentation revealed a total of 92 CNVs (Appendix 2E). Of these CNVs, 64 were found in either the DGV or in the 'structural variation tracks' archived in the UCSC Genome Browser (Appendix 2E and Figure 2.1). These 64 CNVs therefore constitute known copy number polymorphisms (CNPs) and are most likely not associated with the disorder in these MR patients.

The remaining 28 CNVs had not been previously described in any of the structural variation databases, suggesting they constitute either novel CNPs, or rare variants which could contribute to MR pathogenesis. Of these 28 variations, ten autosomal CNVs were excluded from further investigation as they were detected in X-linked families and are thus unlikely to be the direct disease-causing mutation. Also, five CNVs were excluded as they were located in chromosomal regions which, at present, contain no known genes. Ultimately, in total, validation experiments by genomic qPCR were performed for 13 CNVs (Table 2.2).

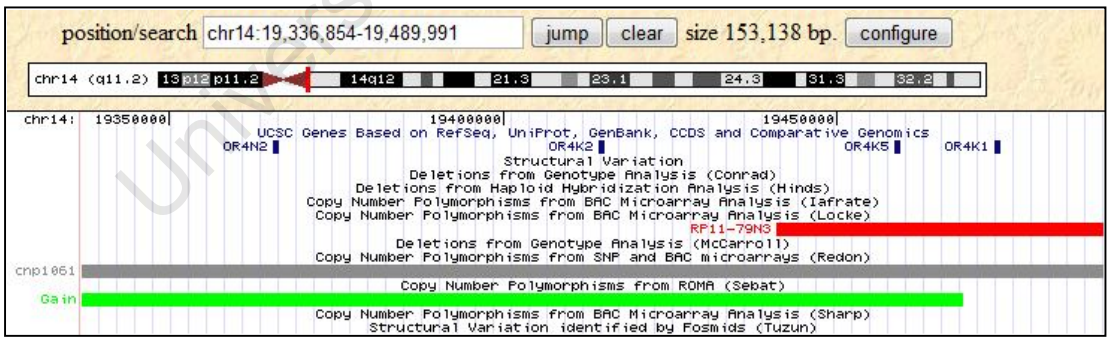


Figure 2.1 An example of a commonly detected CNP. This UCSC Genome Browser exert shows a frequent centromeric 14q11.2 CNP which encompasses several genes involved in olfactory system functioning (*ORN4, ORK2, ORK5, ORK1*). Deletions are illustrate in red (e.g. the 'Locke' track), duplications in green (e.g. the 'Sebat' track) and undetermined copy number changes are illustrated in grey ('Redon' track). Both deletions and duplications at this locus were observed in the South African MR cohort investigated here (Appendix 2E).

Table 2.2 The thirteen CNVs detected by microarray analysis selected for validation using qPCR

	PATIENT	CHANGE	SIZE (Mb)
1.	FX151	6q24.3Dup	1.2
2.	FX233	14q32.12Dup	0.4
3.	FX277	Xp11.22Dup	0.7
4.	FX343	17p11.2Del	3.0
5.	FX444	Xq26.3-27.1Del	2.1
6.	FX444	Xq27.2-27.3Del	1.9
7.	FX444	Xq27.3Del	1.5
8.	FX512	Xp22.11Dup	1.4
9.	FX512	Xp11.3Dup	1.7
10.	FX517	Xq25Del	0.1
11.	XMR14	Xq25Del	0.1
12.	XMR8	Xq25Dup	0.8
13.	XMR12	4p16.3Del	1.6

2.3.2 VALIDATION OF CNVS BY GENOMIC qPCR

Of the 13 CNVs subject to validation, six were shown to be false positives:

1. The 0.4Mb 14q32.12Dup in individual Fx233 (Figure 2.2)
2. The 0.7Mb Xp11.22Dup in individual Fx277

3&4. The 1.4Mb Xp22.11Dup and the 1.7Mb Xp11.3Dup in individual Fx 512

5&6. The 0.1Mb Xq25Del identified in individuals Fx517 and XMR14

The remaining seven novel CNVs were confirmed using qPCR (Table 2.3). Of these, two CNVS were causative of the MR phenotype (in individuals Fx343.1 and XMR8.1). Firstly, the 17p11.2Del was identified in individual Fx343.1, this deletion overlaps with the Smith-Magenis microdeletion syndrome. Secondly, a de novo Xq25Dup was detected in individual XMR8.1. The remaining five CNVs are of indeterminate disease-causing nature. The segregation pattern of the 6q24.3Dup and 4p16.3Del identified in individuals Fx151.1 and XMR12.1 could not be determined due to a lack of family DNA material. Finally, the Xq26.3- 27.3Del in family Fx444 has an unknown effect on MR etiology.

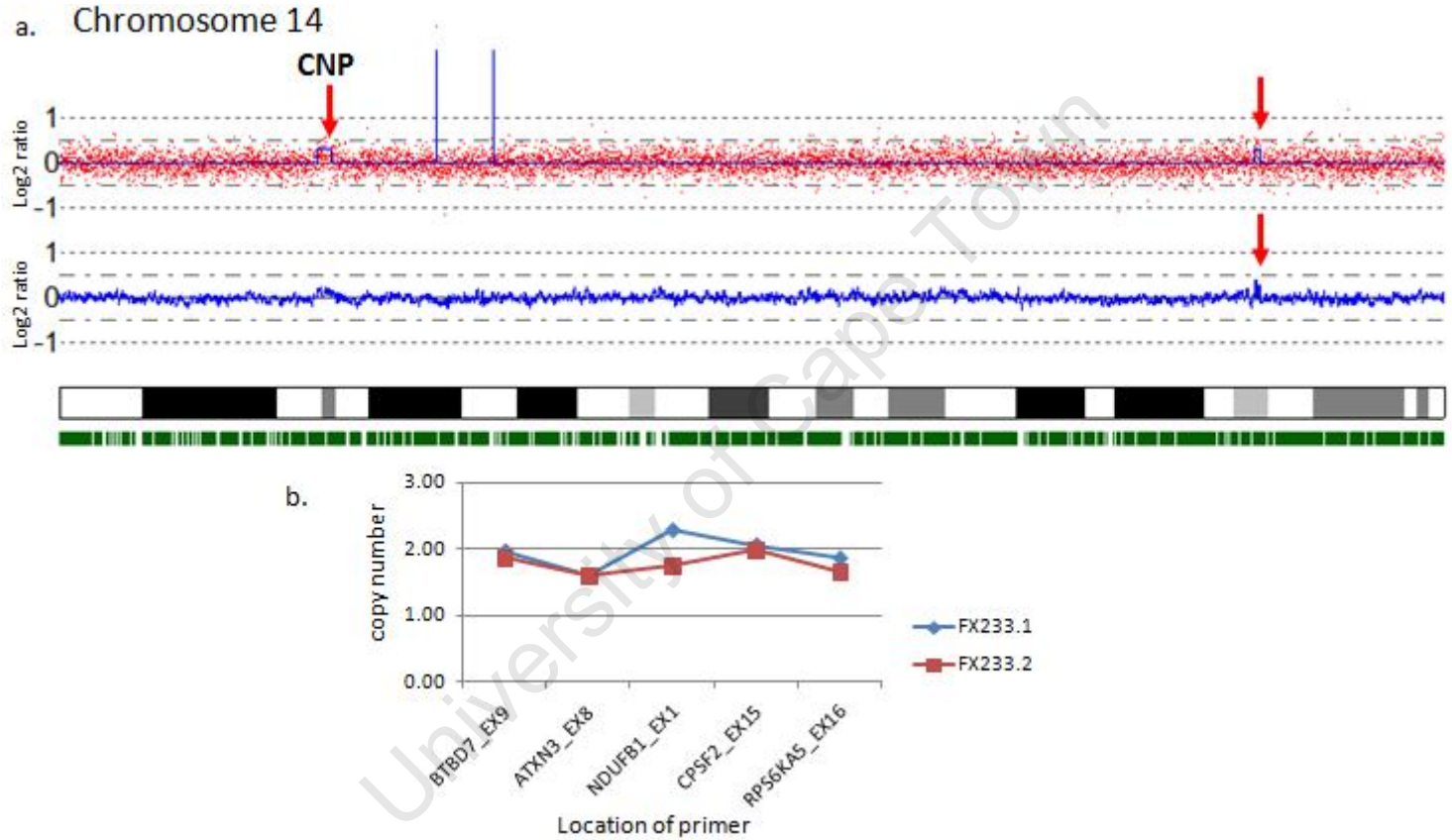


Figure 2.2 An example of a false positive CNV detected by SNP array analysis: a 14q32.12Dup identified in patient Fx233.1. (a) The Log2 ratios of the raw data (upper red track) and the normalised data (lower blue track) for each microarray SNP on chromosome 14. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The 14q32.12Dup reflected positive Log2 ratios and is illustrated by the red arrows. (b) By qPCR this CNV was shown to be a false positive in both affected males (Fx233.1/2), as reflected in the allele copy number of two across the region (i.e. for the external primers designed in *BTBD7* and *RPS6KA5* and internal primers *ATXN3*, *NDUFB1*, *CPSF2*)

Table 2.3 Details of the seven novel CNVs detected by the 250K array and confirmed by genomic qPCR

	PATIENT	CHANGE	SIZE(Mb)	INHERITANCE	CONFIRMED IN FAMILY	GENES ENCOMPASSED	DISEASE-CAUSING	FIGURE
1.	FX151	6q24.3Dup	1.2	XLMR sib-pair	Only DNA from mother available – negative for 6q24.3Dup	*EPM2A* *SNPRH* *FBX030* *GRM1*	undetermined	2.3
2.	FX343	17p11.2Del	3.0	Isolated male	Familial DNA not available	Numerous	Yes Smith-Magenis microdeletion syndrome	2.4
3,4,5.	FX444	Xq26.3- 27.3Del*	~9	XLMR	Both mother and unaffected brother carry the deletion	Numerous see figure 5.5	undetermined	2.5
6.	XMR8	Xq25Dup	0.8	Isolated male	Mother not carrier Duplication arose de novo	*GRIA3* *THOC2* *XIAP* *THOC2*	Yes	2.7
7.	XMR12	4p16.3Del	1.6	Isolated female	Familial DNA not available	At least *ADRAC2C* (see figure 2.9 for more details)	Undetermined Deletion lies at the distal end of the WHS[#] locus. An abnormality on 6p was also previously identified by standard G-band karyotyping	2.8 and 2.9

* Whole-genome CNV analysis suggested three distinct deletions spanning the Xq26.3-27.3 chromosomal region; however, genomic qPCR investigations in this chromosomal region confirmed indicated that the variant was in fact continuous in nature.

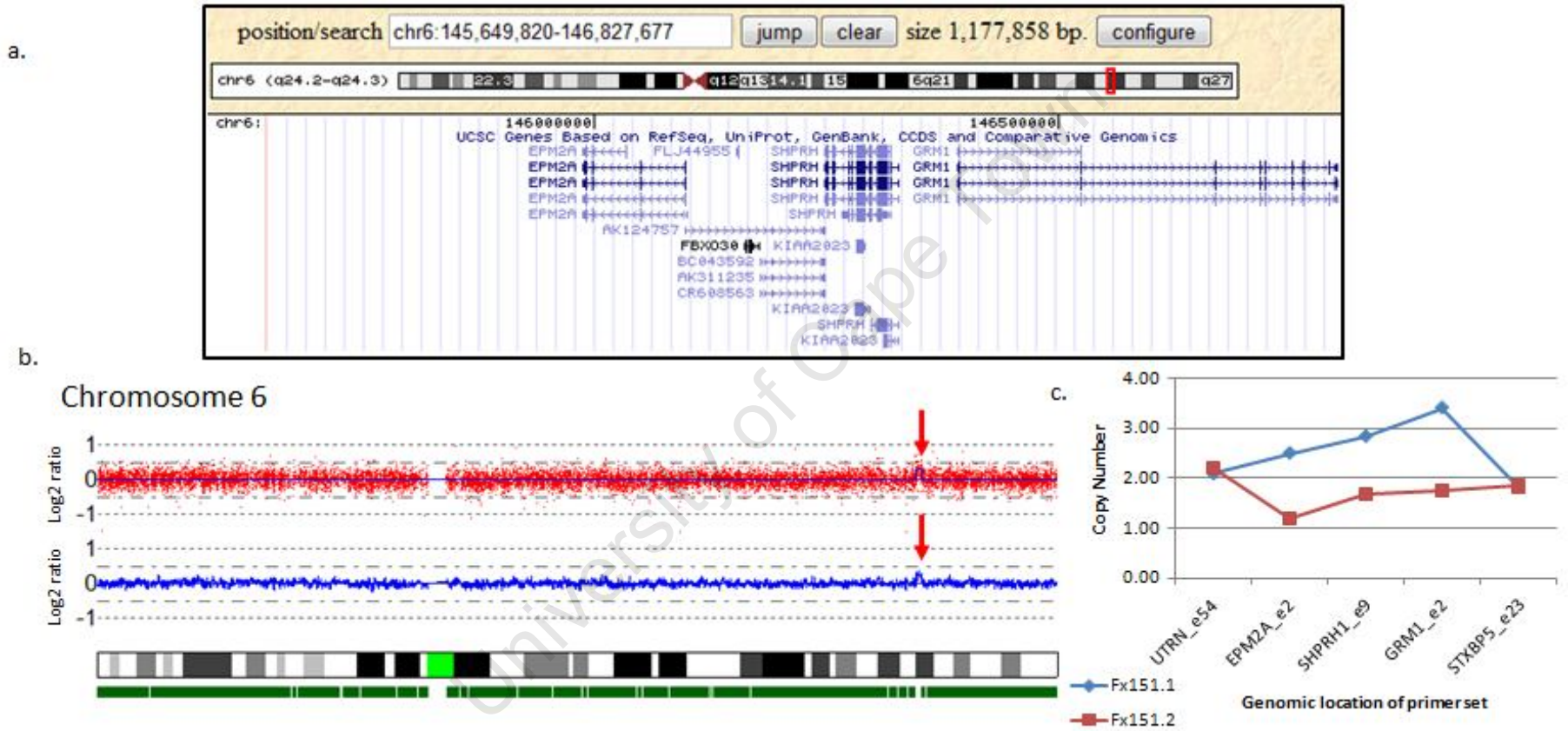


Figure 2.3 The 6q24.3Dup identified in MR patient, Fx151.1 (a) An exert from UCSC illustrating the genes encompassed by the 6q24.3Dup (b) The Log2 ratios of the raw data (upper red track) and the normalised data (lower blue track) for each microarray SNP on chromosome 6. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The 6q24.3Dup reflected positive Log2 ratios and is illustrated by the red arrows. (c) By qPCR, it was concluded that this CNV includes the chromosomal region encompassing *EPM2A*, *SHPRH1* and *GRM1* as indicated by the copy number of three in patient Fx151.1 as compared to the normal autosomal copy number (two) at the flanking primer sets (*UTRN*, *STXBP5*). By comparison, the individuals mother (Fx151.2) does not appear to be a carrier of this CNV, and can be argued for since the difference in copy number between patient and mother was approximately one throughout the primer pairs.

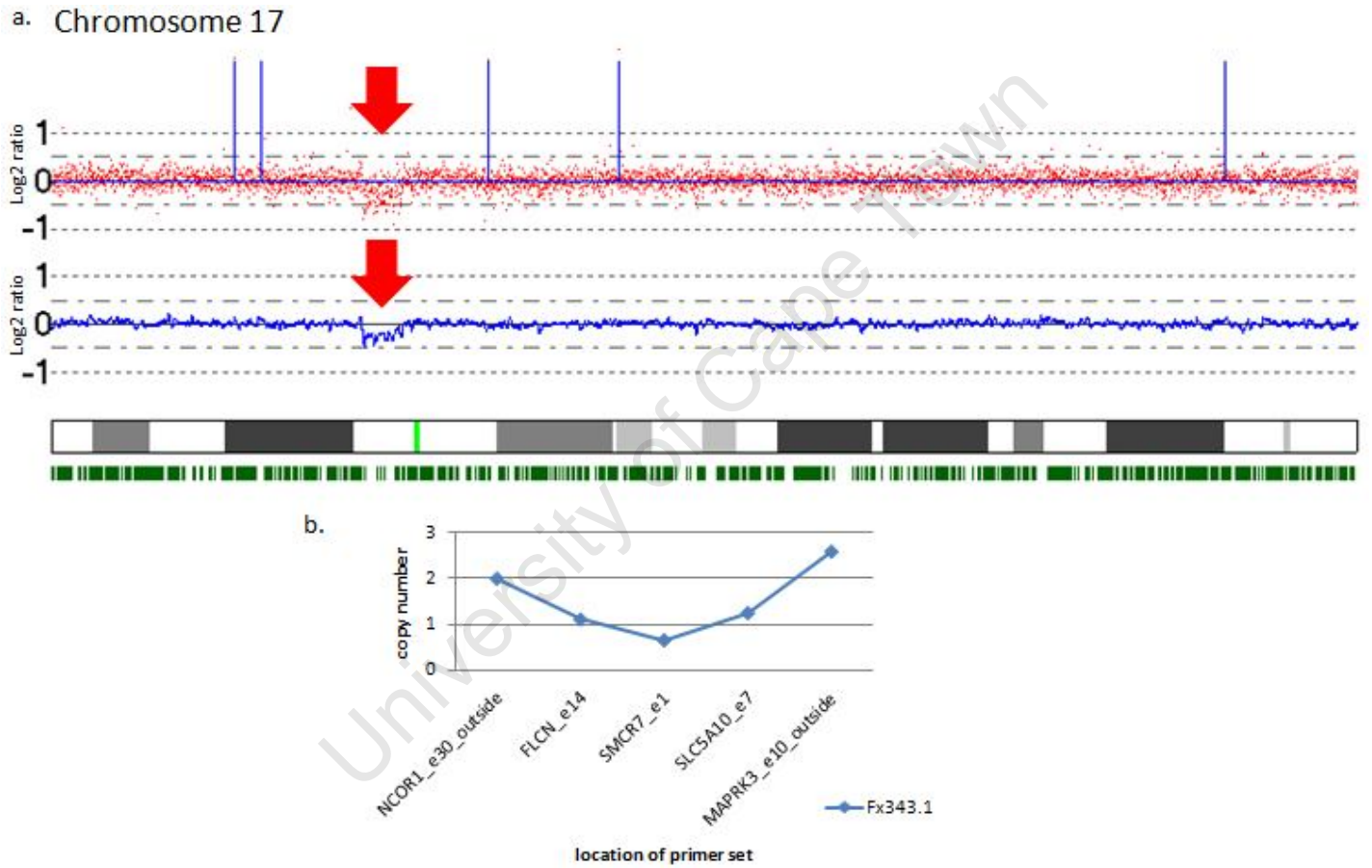


Figure 2.4 The 17p11.2 Del in patient, Fx343.3 causing Smith-Magenis Syndrome. (a) The Log2 ratios of the raw data (upper red track) and the normalised data (lower blue track) for each microarray SNP on chromosome 17. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The 17p11.2Del reflected negative Log2 ratios and is illustrated by the red arrows. (b) By qPCR, it was demonstrated that this CNV encompasses the genes, *FLCN4*, *SMCR7* and *SLC5A10*, as indicated by the single copy number in patient Fx343.1 as compared to the normal autosomal copy number (two) at the flanking primer sets (*NCOR1*, *MAPRK3*).

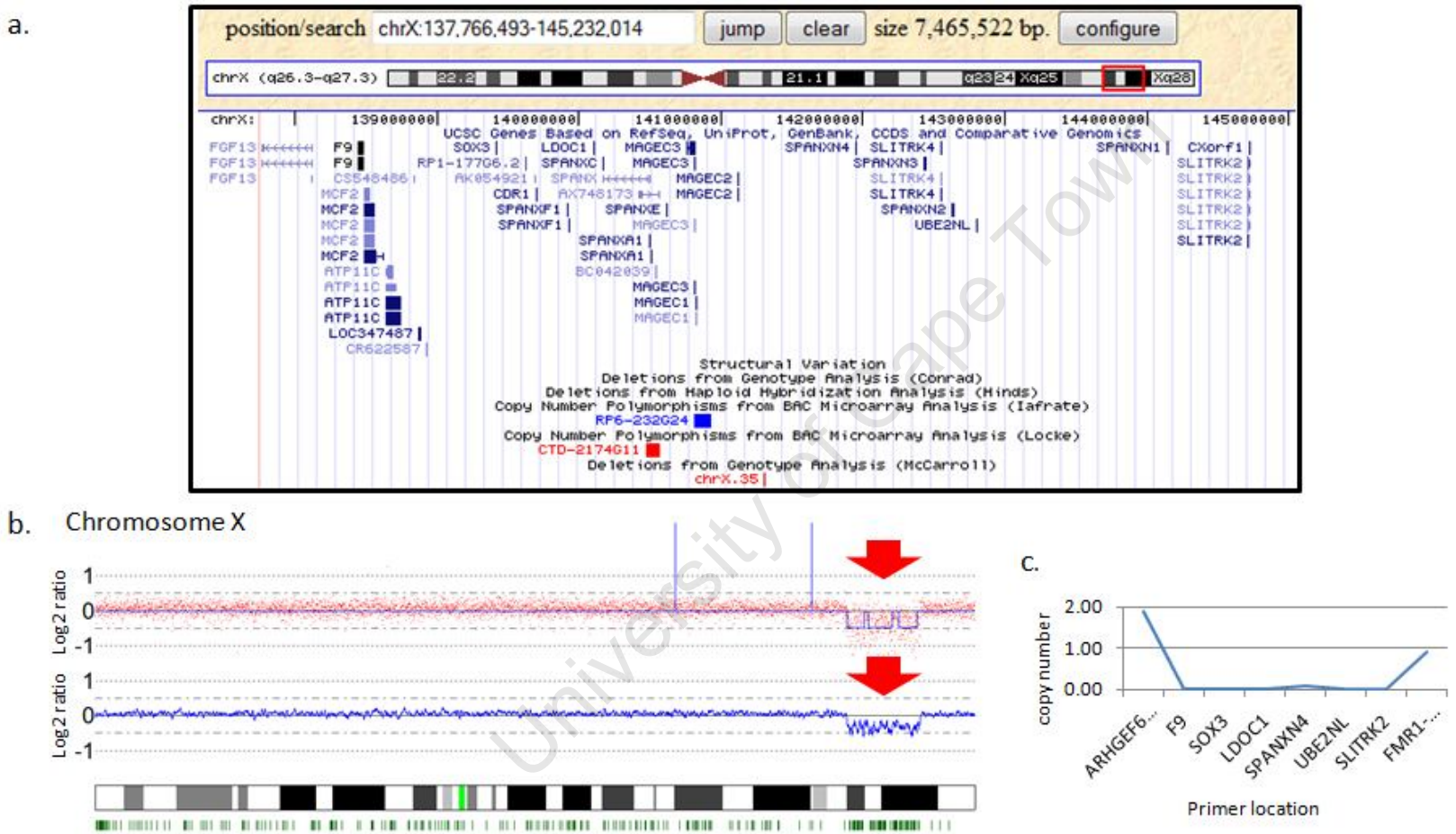


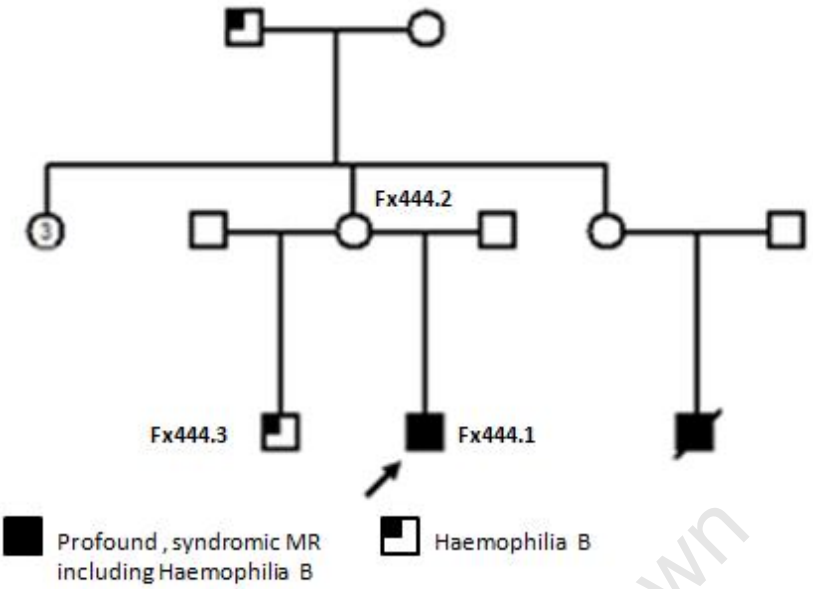
Figure 2.5 The Xq26.3-27.3Del in patient Fx444.1 (a) An exert from UCSC illustrating the genes and known CNVs encompassed by the Xq26.3-27.3Del (b) The Log2 ratios of the raw data (red track) and the normalised data (blue track) for each microarray SNP on chromosome X. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The Xq26.3-27.3Del reflected negative Log2 ratios (red arrows). (c) By qPCR this CNV was shown to span the chromosomal region encompassing the *F9* to *SLITRK2* genes as indicated by the copy number of zero (male patient Fx444.1) as compared to the normal male X-chromosomal copy number (one) at the flanking primer sets (*ARHGEF6*, *FMR1*).

Individual Fx444.1 presented with profound MR and a multitude of additional clinical features including: Haemophilia B, growth retardation, Lennox Gestualt epilepsy syndrome, short broad feet and hands with clinodactyly, undescended testes, hypotonia and no speech or ambulation. In total the deletion spans approximately 9Mb and encompasses 21 known genes and eight predicted genes or open reading frames (ORFs) (Figure 2.5a). Not surprisingly, the *F9* gene, mutations in which cause Haemophilia B, was deleted in affected individual Fx444.1 who presented with this clinical feature.

The family history of Fx444 was particularly interesting (Figure 2.6a), with the proband's maternal male cousin presenting with a very similar phenotype (severe MR, seizures, haemophilia B, growth retardation, deceased in pre-teen years). In contrast, the proband's half-brother (Fx444.3) and maternal grandfather presented only with Haemophilia B (the half-brother also has learning disabilities, the cognitive function of the maternal grandfather was not known). DNA samples were only available for the proband's mother (Fx444.2) and half-brother (Fx444.3). Intriguingly, by qPCR it was established that not only was the proband's mother a carrier for this mutation but also that his half-brother inherited the same 9Mb Xq26.3-27.3Del (Figure 2.6b).

A repeated qPCR experiment on fresh DNA samples from all family members yielded the same inheritance pattern of the deletion for the half-brothers. In addition, the deletion breakpoints were defined by qPCR to a proximal chromosomal region of ~900kb and a distal ~500kb region. These findings revealed identical Xq26.3-27.3Del breakpoints in the half-brothers (Fx444.1/3) at this resolution.

a.



b.

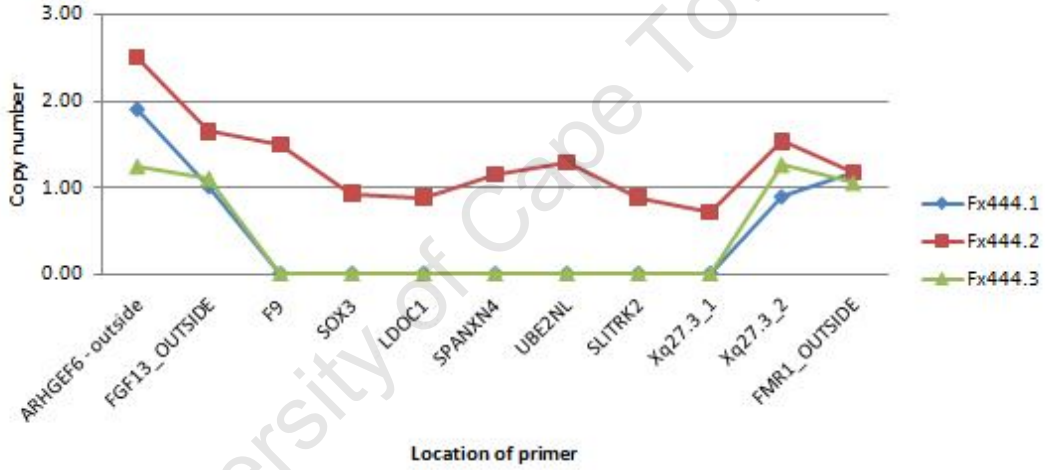


Figure 2.6 The segregation of the Xq26.3-27.3Del in family Fx444. (a) The pedigree of family Fx444. (b) the qPCR profiles demonstrated that the breakpoints of the deletion lie between *FGF13* and *F9* (centromeric break point) and in the gene poor region between genes *SLITRK2* and *FMR1* (telomeric break point). Also, the qPCR results demonstrated that the half-brothers (Fx444.1/3) carry the same deletion and that they inherited this mutation from their carrier mother (Fx444.2).

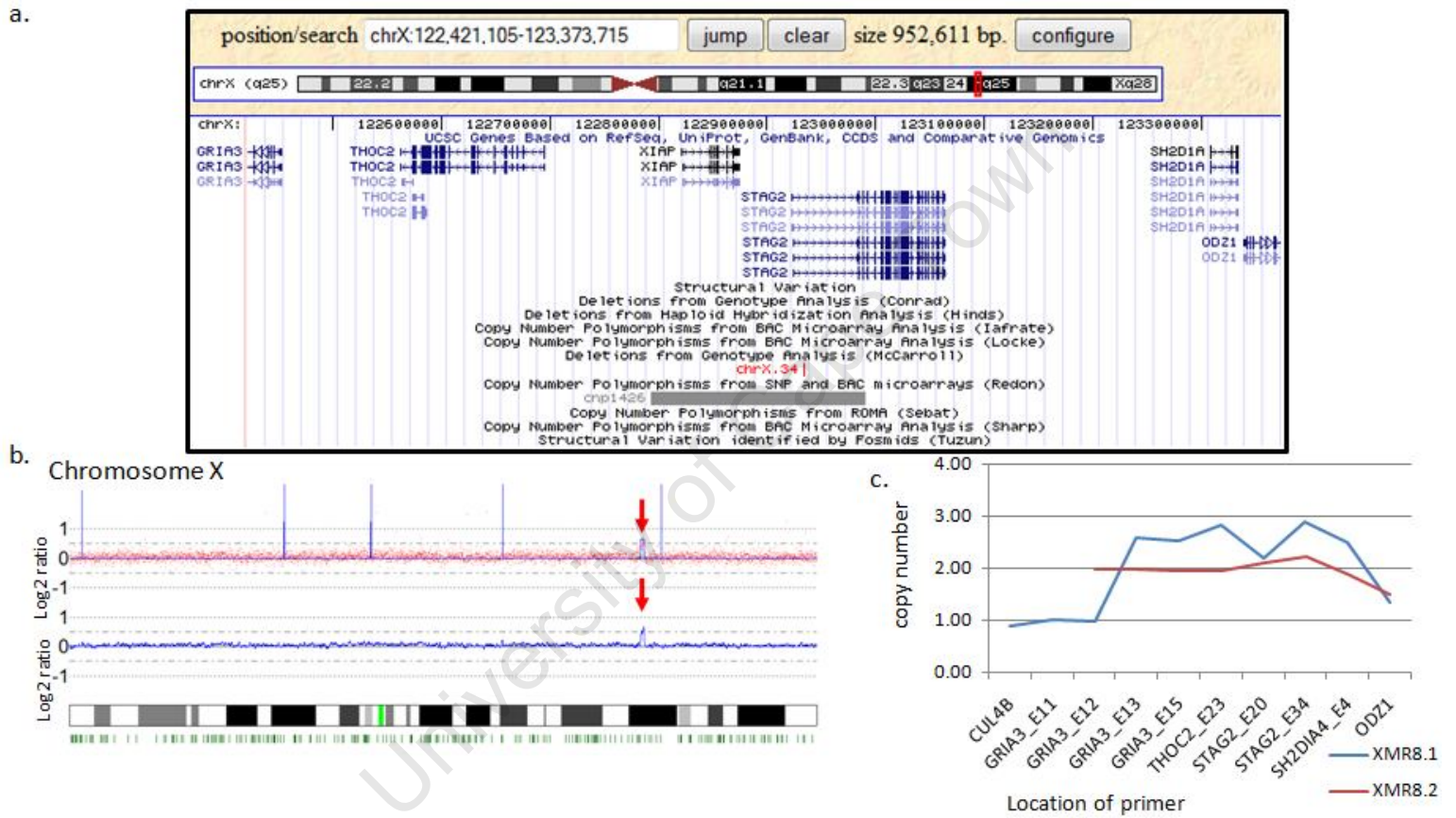


Figure 2.7 The Xq25Dup in patient XMR8.1 (a) An exert from UCSC illustrating the genes and known CNVs encompassed by the Xq25Dup (b) The Log2 ratios of the raw data (upper red track) and the normalised data (lower blue track) for each microarray SNP on chromosome X. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The Xq25Dup reflected positive Log2 ratios and is illustrated by the red arrows. (c) By qPCR, the breakpoints of this CNV were mapped to *GRIA3* intron 11 centromerically and telomerically to lie between *SH2DIA4* and *ODZ1*. The copy number from the qPCR lies between two and three for all 'duplicated probes', it was difficult to determine whether this alteration was a duplication or triplication, however, based on the Log2 ratios the former was more likely. The mother (XMR8.2) of this patient was not a carrier of Xq25Dup (normal copy number of two), the mutation therefore arose de novo in XMR8.1.

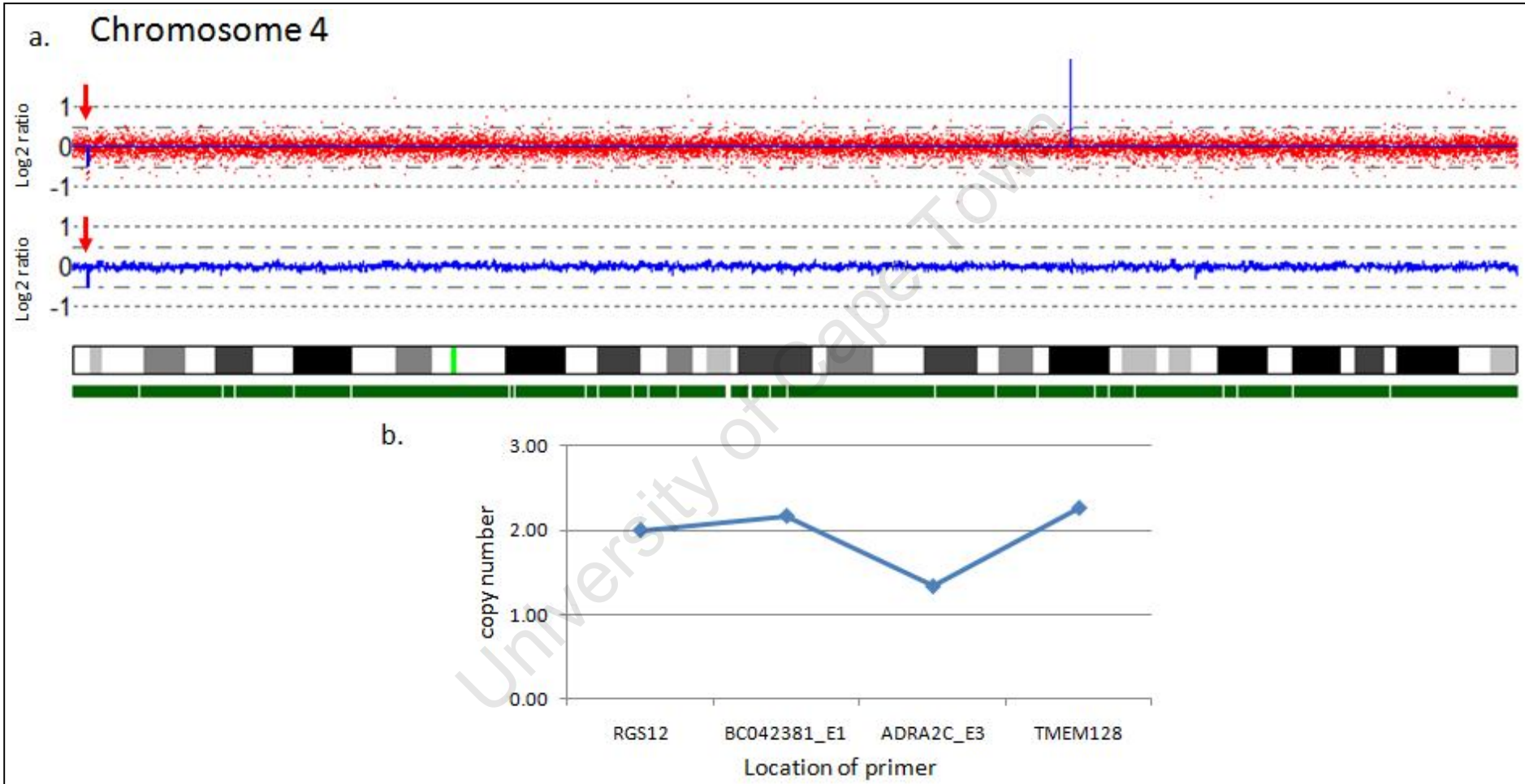


Figure 2.8 The 4p16.3Del in patient XMR12 lies distal to the Wolf-Hirschhorn microdeletion syndrome chromosomal region (a) The Log2 ratios of the raw data (red track) and the normalised data (blue track) for each microarray SNP on chromosome X. Below this data the ideogram is depicted along with the distribution of the SNPS, which appear in green. The 4p16.3Del reflects negative Log2 ratios (red arrows). (b) By qPCR this CNV was shown to include *ADRAC2* as indicated by the copy number of approximately one as compared to the normal autosomal copy number (two) at the flanking primer sets (*BC042381, TMEM128*).

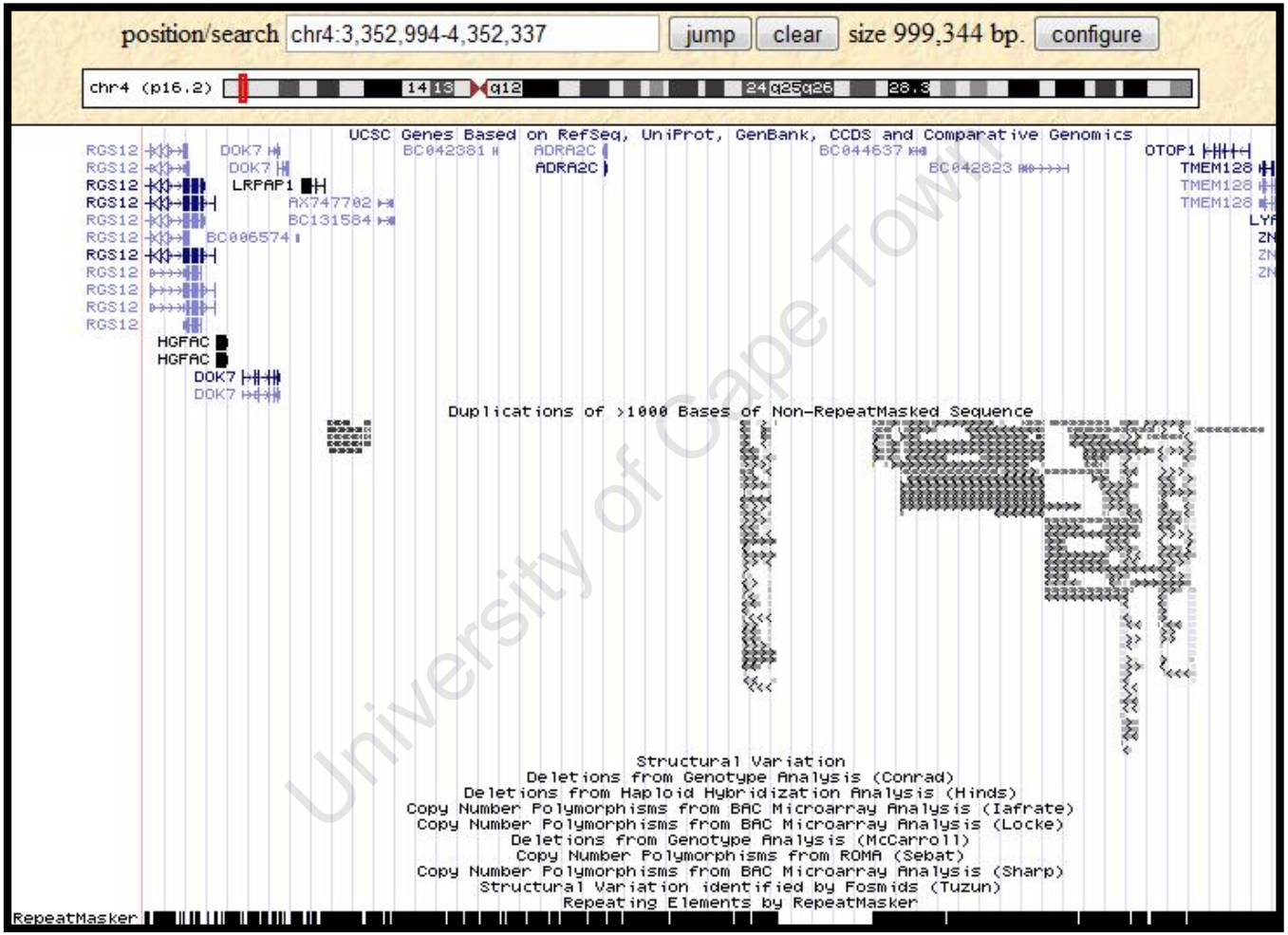


Figure 2.9 An exert from the UCSC Genome Browser illustrating the 4p16.3Del detected in patient XMR12. The region was characterised by a large number of segmental duplications (see 'Duplications of >1000 bases of non-repeat masked sequence' track) and a high frequency of repetitive sequence (see 'Repeat Masker' track). These two genomic architectural traits made primer design required for follow up qPCR virtually impossible.

2.3.3 MOLECULAR FOLLOW-UP INVESTIGATIONS IN FAMILY FX444

2.3.3.1 X chromosome microsatellite marker genotyping in the family trio

In order to further validate that the Xq26.3-27.3Del segregated in the family as suggested by the qPCR results, whole X-chromosome allele segregation analysis was performed using microsatellite markers. This haplotype analysis demonstrated that both the individual with severe syndromic MR (Fx444.1) and his mildly affected half-brother (Fx444.3) inherited the same maternal X chromosome in the chromosomal region encompassing the deletion (Figure 2.10).

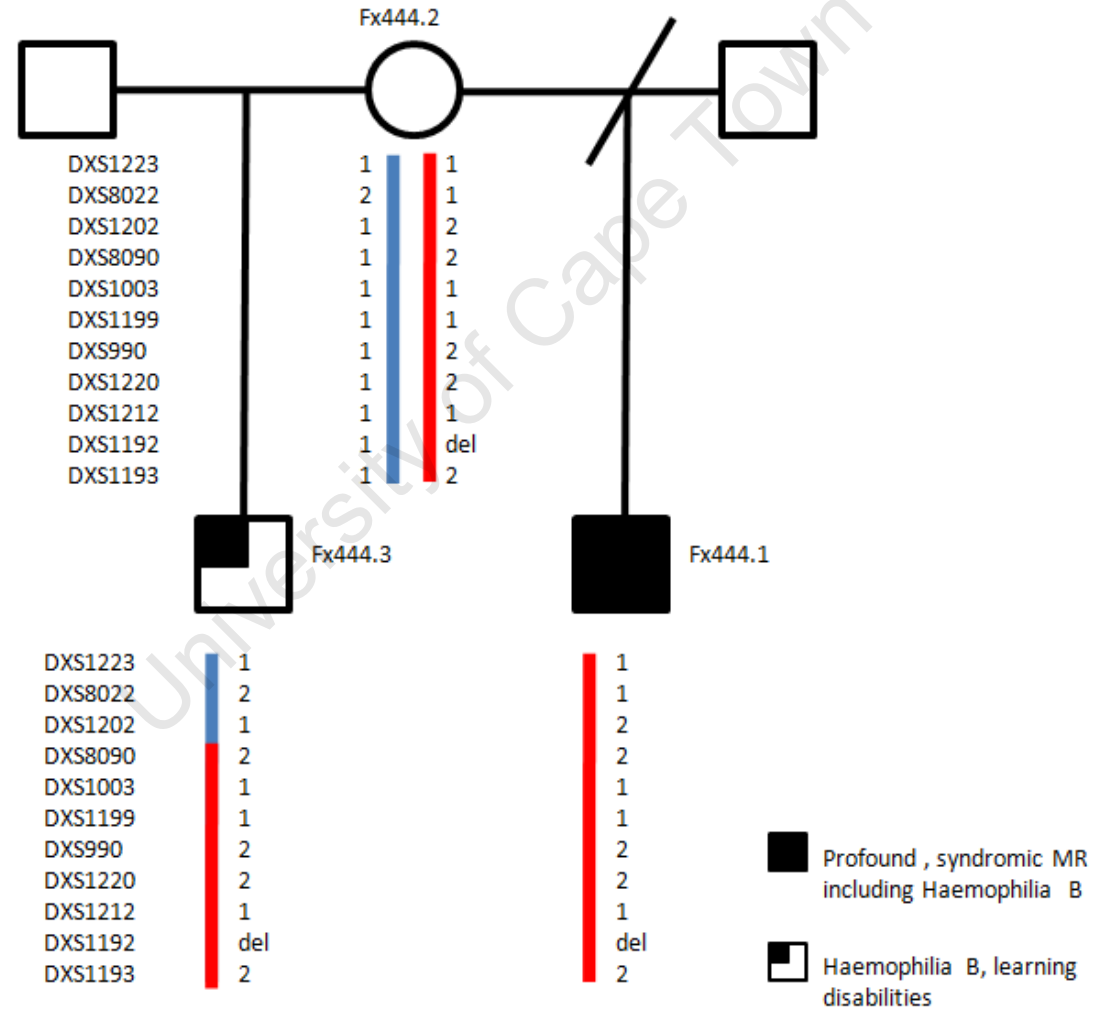


Figure 2.10 The X-chromosomal haplotype segregation in family Fx444. The allele segregation analysis conducted in this family demonstrated that the half-brothers (Fx444.1/3) have inherited the same maternal X chromosome (red bar) in the region of the Xq26.1-27.3 Del (which falls between markers DXS1212 and DXS1193, hence the deletion of marker DXS1192). In fact, the only difference in X-chromosomal genetic material between the half-brothers is owing to a recombination between markers DXS1202 (~26.2Mb) and DXS8090 (~36.7Mb). This haplotype analysis further validated the presence of the Xq26.1-27.3 Del in the half-brothers.

2.3.3.2 X-inactivation analysis in unaffected female Fx444.2

In order to ascertain the X-inactivation pattern in the half-brothers' mother (Fx444.2), X-inactivation analysis was performed at the *MAOA* VNTR locus. From these analyses it was established that the X-inactivation ratio was 45:55 (Figure 2.11). This X-inactivation ratio suggests random X-inactivation in the unaffected female (Fx444.2) with no bias toward preferential inactivation of the X-chromosome with the Xq26.3-27.3Del (i.e. the 318bp *MAOA* VNTR allele)

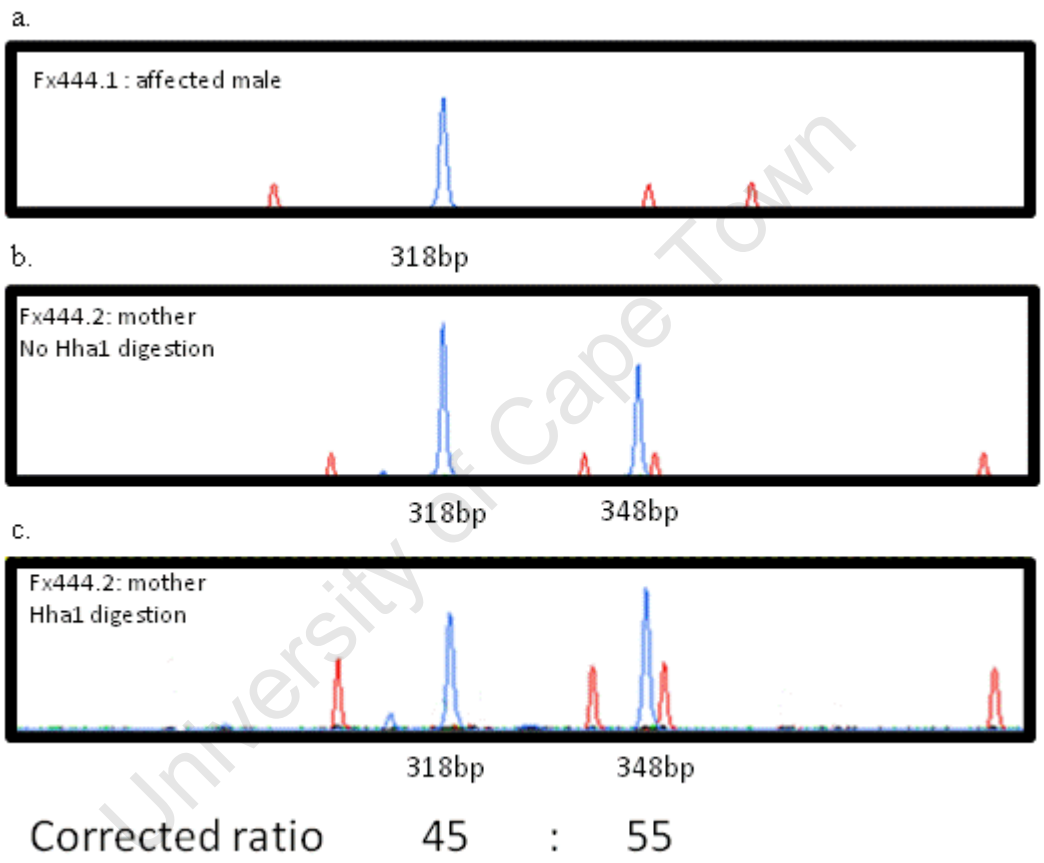


Figure 2.11 The X-inactivation pattern in the unaffected mother (Fx444.2) of the half-brothers Fx444.1 and Fx444.3. (a) The electropherogram generated by size-based electrophoresis illustrating the *MAOA* VNTR allele in affected individual Fx444.1. Both half-brothers carried the 318bp allele (blue) at this locus. The 500bp Rox size standard [Applied Biosytems] is illustrated in red (b) The electropherogram illustrating the intensity of the two *MOAO* VNTR alleles (318bp and 348bp) (blue) prior to digestion of the unaffected mother's DNA sample with the methylation-specific enzyme HhaI. There was evidence in this electropherogram of preferential amplification of the larger (348bp) allele (blue) as compared to the smaller (318bp) allele (blue). (c) The electropherogram illustrating the intensity of the two *MOAO* VNTR alleles (318bp and 348bp) (blue) after digestion of the unaffected mother's DNA sample with the methylation-specific enzyme HhaI. Once the allele-specific PCR amplification bias was accounted for it was concluded that the X-inactivation ratio was 45:55, illustrating a near-random distribution of X-inactivation.

2.4 DISCUSSION

The whole-genome copy number analysis of 30 MR patients with MR and dysmorphism and/or additional clinical features using a 250K SNP array revealed a total of 92 CNVs. Of these 92 CNVs, 79 variants (86%) were excluded from follow-up investigations. A total of 64 (69%) CNVs encompassed known CNPs. Also, the chromosomal location of 10 CNVs (11%) did not concur with familial inheritance pattern and therefore are most likely not the disease-causing mutation. Finally, five CNVs (6%) were predicted to span chromosomal regions harbouring no known genes. While putative effects of these CNVs on the regulatory regions of neighbouring genes cannot be excluded, these variants were not further investigated in this particular study. The exclusion of 89% of the CNVs detected illustrates the structural variability of the genome as described previously (Alkan et al. 2009, Conrad et al. 2009, Redon et al. 2006, Sharp, Cheng and Eichler. 2006). Also, the vast majority of CNVs detected in this study were excluded, thus illustrating that the adopted exclusion criteria, while not optimal, were still effective in interpreting the pathogenic nature of CNVs.

The remaining 13 CNVs (14%) withstood the initial exclusion criteria and were therefore subject to validation by genomic qPCR. Of these, six (46%) variants were shown to be false positives. This result illustrates two important points; firstly the importance of verifying all CNVs detected by microarrays using a second quantitative technique. Secondly, the statistical modelling used to predict copy number was evidently sub-optimal. In this study CNAG was used to predict copy number from the genotype calls generated by the Affymetrix GTYPE software. The CNAG software calculates log2 ratios from the genotype call data of both the test and reference sample/s and subsequently uses a hidden markov model (HMM) to generate allele calls. This program has certain weaknesses best illustrated by the following examples. Firstly, all false positives generated the expected allele calls of three or one (the designated values for duplications and deletions respectively) using the HMM. Secondly, all false positive changes showed the deletion of at least four consecutive SNPs, or the duplication of six consecutive SNPs, as recommended previously (Hehir-Kwa et al. 2007). These observations suggest that the selected

thresholds, both in terms of number and intensity, of duplication/deletion calls was insufficiently stringent. However, the converse was also shown to be true as the HMM failed to predict allelic loss for the 17p11.2Del identified in individual Fx343.1. This deletion was only detected by manual Excel visualisation of the Log2 ratios and verified by qPCR. These data clearly illustrate that the algorithms used to calculate copy number are not optimal. It should be noted that other platforms which use two-flurophore hybridisation methods (where test and reference sample are competitively hybridised to and array as oppose to the in silico analysis adopted here) suffer from similar sensitivity problems. It was concluded that the CNAG algorithm is useful in automatically identifying potential CNVs, however, validation by a second technique is recommended, as is manual assessment of data.

The remaining seven CNVs were shown to constitute 'real' copy number changes, i.e. identified by the SNP array and verified using qPCR. These seven CNVs included a duplication of indeterminate disease-association (6q24.3Dup) in family Fx151.1, as well as one known disease-causing alteration (17p11.2Del) in individual Fx343.1. Most interestingly, a novel 9Mb X-linked deletion (detected as three separate deletions on the SNP array) (Xq26.3-27.3Del) was detected in a severely affected MR patient as well as his 'unaffected' brother, suggesting this variant exhibits low penetrance. Also, a novel de novo disease-causing mutation (Xq25Dup) in XMR8 was detected. Finally, a 4p16.3Del that falls within a known microdeletion syndrome chromosomal region (Wolf-Hirschhorn syndrome) was identified in individual XMR12.1, this patient was also shown to carry a chromosome 6p abnormality by G-banded karyotype. The disease-causing nature of these alterations in this patient are unknown at this stage.

The approximately 1.2Mb 6q24.3 duplication identified in patient Fx151.1 was not inherited from the unaffected mother (Fx151.2). Further attempts to assess the segregation of this variant in the family were thwarted by the inability to contact additional family members (affected male sib and unaffected father). Therefore, given this lack of additional familial DNA material, the pathogenic nature of this variant remains to be established. However, it was interesting to note that the duplicated chromosomal region encompasses *EPM2A* and *GRM1*, which could pose as MR candidate genes. Firstly, mutations in *EPM2A* have been previously identified

in patients with a progressive epilepsy syndrome, myoclonic epilepsy of Lafora (OMIM **#254780**). While the patients in family Fx151 do not present with epilepsy, given the role of *EPM2A* in cognitive function, this gene presents an interesting pathogenic MR candidate gene. Also, the duplicated gene, *GRM1* functions as a glutamate receptor, these receptors are involved in glutamatergic neurotransmission. There are numerous examples of defects in the glutamatergic neurotransmission system causing cognitive defects. For instance, mutations in *GRIA3* (also a glutamate receptor) have been shown to cause XLMR (Wu et al. 2007). Also, excessive stimulation of mGluR5 (metabotropic glutamate receptor 5) was demonstrated to be an important pathogenic mechanism underlying the MR etiology in Fragile X syndrome patients (Antar et al. 2004). Finally, CNVs encompassing genes that comprise the glutamate receptor complex are enriched in patients with MR (Poot et al. 2009). Collectively, this evidence suggests that dosage effects of *EPM2A* and *GRM1* may contribute to the disease aetiology in family Fx151.1, however, no conclusions can be reached until the segregation of this CNV in the family has been established.

The 3Mb, 17p11.2Del detected in patient Fx343.1 falls within the pericentromeric chromosomal region in which heterozygous deletions cause Smith Magenis syndrome. This syndrome is characterised by MR, a typical facial dysmorphism, behavioural problems, cardiac and genitourinary abnormalities (OMIM **#189920**). Individual Fx343.1 presented with a syndromic form of MR including hyperphagia, behavioural problems, micropenis and a high pain tolerance. Therefore, the clinical presentation of this patient overlapped with the some of the features of Smith-Magenis syndrome. This prompted, a detailed retrospective review of this patient's medical records (which were not available prior to this study) which showed that a diagnosis of Smith-Magenis syndrome was previously made on the basis of a positive FISH result at 17p11.2. Therefore, it was concluded that this 3Mb 17p11.2del was the MR-causing mutation in Fx343.1 and that this deletion was most likely de novo in origin (given the normal cognitive function in this individual's parents).

Individual Fx444.1 was found to carry an approximately 9Mb X chromosomal deletion (Xq26.3-27.3Del) encompassing 21 known genes (Table 2.4). This deletion was transmitted from the carrier mother (Fx444.2) to both her sons (who have different fathers), Fx444.1 and Fx444.3 as confirmed by both qPCR and haplotype analysis. Interestingly, despite the half-brothers carrying the same 9Mb deletion, their phenotypic expression was highly variable. The severely affected proband (Fx444.1) presents with profound MR, Haemophilia B, growth retardation, Lennox Gestualt epilepsy syndrome, short and broad feet and hands with clinodactyly, undescended testes, hypotonia and an absence of speech or ambulation. Conversely, the proband's half-brother (Fx444.3) presents with haemophilia B and learning disabilities with no additional clinical features.

Equally surprising was the fact that the chromosomal region in which the 9Mb Xq26.3-27.3Del lies was not a gene-poor interval and encompasses a total of 21 known genes and eight uncharacterised transcripts. Male X-chromosomal hemizygosity renders the Xq26.3-27.3Del males in this family null for the proteins encoded by the 21 known genes in this region. Thus the haemophilia B phenotype of the half-brothers can be accounted for by the deletion of *F9* which encodes blood clotting factor IV.

Similar large Xq deletions are reported in the literature in patients with syndromic MR (Parvari et al. 1999, Rousseau et al. 1991). However, in contrast to the deletion described here, all reported deletions encompass either the known MR genes, *FMR1* (distal to this deletion) or *PHF6* (proximal to this deletion). The deletion of these two genes was reported to cause the MR phenotype in these XqDel patients. Conversely, given that the Xq26.3-27.3Del mutation described here does not encompass these MR genes, it was difficult to draw parallels between the phenotype in these patients and the affected individuals of family Fx444

Table 2.4 The genes encompassed within the Xq26.3-27.1Del chromosomal region identified in Family Fx444

	Gene	Associated disorder	Function
1.	*F9*	Haemophilia B (OMIM #306900)	Factor IV clotting factor
2.	*MCF2*	Loss of the proteins N-terminal owing to recombination gives the mutated protein oncogeneic potential	GDP-GTP exchange factor that modulates GTPase activity of Rho family
3.	*ATP11C*	None known (another ATPase family member (*ATP6AP2*) has been previously implicated in XLMR) (Ramser et al. 2005)	ATPase, Class VI, type 11C
4.	*SOX3*	XLMR with growth hormone deficiency (Laumonnier et al. 2002)	Transcription factor involved in embryological development and determination of cell fate
5.	*CDR1*	target molecule of autoantibodies in patients with paraneoplastic cerebellar degeneration	Cerebellar degeneration-related protein 1
6.	*LDOC*	Putative tumour suppressor gene	Putative regulator of the transcriptional response mediated by nuclear factor kappa B (NF-KappaB)
7.	*SPANX*	None known	The *SPANX* genes are a group of sperm associated proteins which encode differentially-expressed testes-specific proteins that localise to various sub-cellular compartments.
8.	*SPANXA1*		
9.	*SPANXC*		
10.	*SPANXE*		
11.	*SPANXF1*		
12.	*SPANXN1*		
13.	*SPANXN2*		
14.	*SPANXN3*		
15.	*SPANXN4*		
16.	*MAGEC1*	tumour-specific antigens, putative cancer immunotherapy targets	Members of the melanoma antigen gene (MAGE) family. They are tumour-specific antigens that are recognised by autologous cytolytic T lymphocytes. Not expressed in normal tissue except the testis and various tumours
17.	*MAGEC2*		
18.	*MAGEC3*		
19.	*UBE2NL*	None known, However, an X-chromosomal UBE family member, UBE2A has been implicated in XLMR (Nascimento et al. 2006)	Ubiquitin-conjugating enzyme
20.	*SLITRK2*	None known, however, other family members have been associated with various disorders: *SLITRK1*: Gilles De La Tourette syndrome (GTS)	Members of the SLITRK family are integral membrane proteins, also the proteins' C-terminal shows homology with neurotrophin receptors. These genes are expressed primarily in the neurons and have neurite modulating activity.
21.	*SLITRK4*		

Interestingly, the *SOX3* gene was also encompassed in the Xq26.3-27.3Del. *SOX3* mutations have been shown to cause MR. (Laumonnier et al. 2002). Furthermore, *ATP11C* and *UBE2NL* both belong to protein families previously implicated in MR (*ATP6AP2* and *UBE3A/UBE2A* respectively). Finally, *SLITRK2* and *SLITRK4* are neuronal transmembrane proteins that control neurite outgrowth (Aruga and Mikoshiba. 2003, Aruga, Yokota and Mikoshiba. 2003). SLITRK protein family members are thought to play a role in the disease-mechanisms of associated neurological disorders including Gilles de la Tourette Syndrome (*SLITRK1*, OMIM **#137580**) and psychiatric disorders (*SLITRK2*) (Smith et al. 2009). Collectively, the function of at least some of these 21 deleted genes would appear to account for the MR phenotype in family Fx444. However, given the half-brothers' phenotypic variability the pathogenic role of this deletion was not immediately clear. There are two potential hypotheses for this variability. First, that the Xq26.3-27.3Del is not the disease-causing mutation in this family and that the MR etiology is owing to another mutation/s. Alternatively, the Xq26.3-27.3Del is the disease-causing mutation in this family, but exhibits low penetrance.

The first hypothesis assumes the phenotypic variability observed in the family was owing to an alternative mutation. At this point it was useful to examine the pedigree of family Fx444 (Figure 2.6a). Given the haemophilia B phenotype in the half-brothers' maternal grandfather, it was assumed that the Xq26.3-27.3Del was originally inherited from this individual through the maternal line (Fx444.2). Similarly, it was assumed (but cannot be proven) that the half-brothers' maternal cousin inherited the deletion from the maternal grandfather by transmission through his mother. The proposed unidentified mutation could be inherited in either an X-linked or autosomal fashion, although the pedigree supports an X-linked mode of inheritance. Given the lack of syndromic clinical presentation in the maternal grandfather it was assumed that a second X-linked mutation could not have been inherited from him. Alternatively a second X-linked mutation could have been inherited from the maternal grandmother. A recombination between the maternal and paternal chromosomes in Fx444.2 and her sister could lead to both the unidentified X-linked mutation and the Xq26.3-27.3Del being present on the same allele. This inheritance pattern would account for the syndromic MR presentation and haemophilia B in proband Fx444.1 and his similarly affected maternal cousin.

However, if this were the case it would be expected that the female carriers (Fx444.3 and her sister) would present with at least a mild MR phenotype, yet this clinical feature is not present in these individuals, suggesting an X-linked inheritance of a second mutation unlikely. Second to this, the X-inactivation pattern in the half-brothers mother (Fx444.2) was shown to be random (45:55) by X-inactivation studies at the *MAOA* VNTR locus. Given that up to 50% of XLMR carrier females exhibit skewed X-inactivation (i.e. >>80:20) (Plenge et al. 2002), the random X-inactivation pattern observed in this individual may suggest that the carrier females do not carry a second XLMR mutation. An autosomal mutation is also unlikely given the strong X-linked inheritance pattern in the family and that the severely affected maternal cousins share only 6.25% genetic material.

Low penetrance of Xq26.3-27.3Del can also account for the Fx444 familial phenotypic variability since mouse models of the XLMR gene *SOX3* exhibit a similar phenotypic variability. One third of *Sox3* male knockout mice were phenotypically normal, while the remaining two thirds exhibited a spectrum of abnormalities including craniofacial defects, poor growth and general weakness, the most severely affected mice did not survive to the weaning stage (Figure 5.12) (Rizzoti et al. 2004)

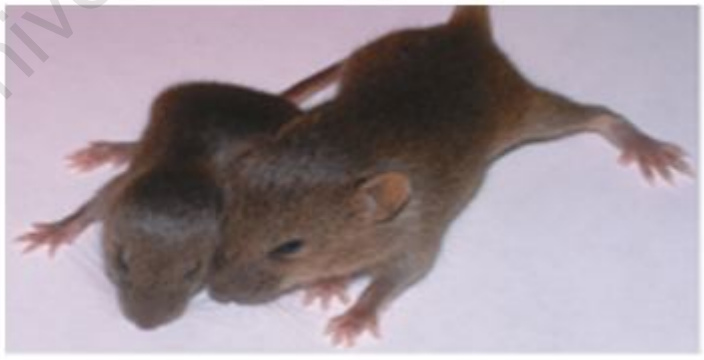


Figure 2.12 A pair of Sox3 null littermates at 2 weeks illustrating the degree of phenotypic variability. Here the phenotypic variability is clearly evident in the level of growth retardation as well as the craniofacial defects (from Rizzotti et al. 2004)

In human studies, a *SOX3* polyalanine expansion mutation has been reported to cause XLMR with growth hormone deficiency (Laumonnier et al. 2002). In addition, a detailed MRI study revealed that *SOX3* mutations resulted in infundibular hypoplasia and hypopituatism in all patients assessed (leading to growth hormone deficiency) but MR was not a consistent feature (Woods et al. 2005). Finally, duplications of the gene have been reported in patients with X-linked hypopituitarism and variable levels of MR (Solomon et al. 2004). These investigations all illustrate the variable level of MR associated with *SOX3* mutations and reflect the phenotypic variability observed in this family. It would therefore be interesting to compare the half-brothers pituitary function at a biochemical level (e.g. luteinizing hormone (LH) and follicle-stimulating hormone (FSH) and thyroid stimulating hormone (TSH) levels) to elucidate any potential differences between the males on a sub-clinical level. However, owing to difficulties contacting family Fx444.1 these investigations have not been possible at present.

These aforementioned studies and observations may explain the observed phenotype of a single *SOX3* deletion. However, it must still be noted that the Xq26.3-27.3Del encompasses an additional 20 genes (*F9* leads to the haemophilic phenotype). Of these genes, nine (*SPANX, SPANXA1, SPANXC, SPANXE1, SPANXF, SPANXN1, SPANXN2, SPANXN3, SPANXN4*) form part of the sperm associated protein family. This family of proteins encode differentially-expressed testes-specific proteins that localise to various sub-cellular compartments and thus are unlikely to contribute to the disease-etiology in the family (Zendman et al. 2003). Similarly, three genes belong to the melanoma antigen gene (MAGE) family which encode tumour specific antigens and are not predicted to play a role in the phenotype of family Fx444. The remaining genes *MCF2, CDR1* and *LDOC* can also be potentially excluded from playing a role in the syndromic MR pathogenesis given the putative role of these genes at various stages of cancer progression.

The remaining deleted genes *ATP11C, UBE2NL, SLITRK2, SLITRK4* could be hypothesised to contribute to the familial MR etiology. *ATP11C* and *UBE2NL* both belong to protein families which have been implicated in MR (Nascimento et al. 2006, Ramser et al. 2005). Similarly, the *SLITRK* protein family have been shown to play a role in neuronal outgrowth and have been associated with other neurological

disorders (Aruga and Mikoshiba. 2003, Aruga, Yokota and Mikoshiba. 2003, Smith et al. 2009). However, all four of these genes form part of protein families whose members may have overlapping functions. This redundant function may account for the lack of a severe syndromic MR phenotype in some Fx444 family members. Ultimately the Xq26.3-27.3Del phenotype may be modified by genetic variants in not only the protein families to which *ATP11C*, *UBE2NL*, *SLITRK2/4* and *SOX3* belong, but also by genetic and epigenetic variation acting on the biological systems implicated in the pathogenesis of syndromic MR in this family.

In conclusion, we hypothesise that the Xq26.3-27.3Del is most likely the disease-causing mutation in family Fx444.1, although a second mutation cannot be excluded. This hypothesis was supported by the phenotypic variability observed in *Sox3* knock-out mice, as well as the variability of clinical presentation (particularly with respect to MR) in patients with *SOX3* mutations (Rizzoti et al. 2004, Solomon et al. 2004, Woods et al. 2005). Furthermore, inter and intra-familial variability for CNVs has been reported before, with individuals carrying a 15q13 dup/del (van Bon et al. 2009) and 16p11.2Del (Bijlsma et al. 2009) presenting with variable phenotypes. However, this is the first report of extreme intra-familial variability in hemizygous males for an X-linked deletion.

In an isolated case of MR, patient XMR8 was found to carry a de novo 0.8Mb Xq25Dup which encompassed four known genes: *GRIA3, STAG2, THOC2* and *XIAP*. The duplication breakpoints were determined to lie within *GRIA3* intron 12 (proximal end) and in the intragenic region between genes *STAG2* and *ODZ1* (proximal end).

Similar Xq25 duplications have been reported in two patients with MR, the dosage-sensitive gene within the duplicated region was concluded to be *GRIA3* (Bonnet et al. 2009, Chiyonobu et al. 2007) (Table 2.5). Mutations which affect GRIA3 (including missense mutations, gene deletions) function have been previously reported in patients with mild-moderate NS-XLMR (Wu et al. 2007). Similarly, the previously described Xq25Dups resulted in reduced levels of *GRIA3* in one patient (Chiyonobu et al. 2007), and aberrant *GRIA3* transcripts in the second patient (Bonnet et al. 2009). These duplications encompass exons within the 5'end of the *GRIA3*

transcript. However, in contrast the Xq25Dup in XMR8 identified in this study, involved duplication of the last four exons of this transcript. Therefore, essentially it was predicted that this patient may still have one functional copy of the *GRIA3* transcript. Unfortunately, mRNA could not be obtained from this patient to confirm the presence of normal levels of the compleete *GRIA3* transcript. Thus it could not be ruled out that the duplication affects the *GRIA3* transcript in a manner similar to the previously described duplications.

Table 2.5 Clinical and molecular features of patients carrying Xq25Dup

MUTATION	GENES ENCOMPASSED	PHENOTYPE	REFERENCE
Xq25Dup	*GRIA3* (partial- first 6 exons)	Male patient presenting with: Severe MR autism short stature minor facial dysmorphism (hypertelorism, epicanthic folds, short neck)	(Chiyonobu et al. 2007)
Xq25Dup Two small duplications	Duplication 1 *GRIA3* (partial- first 11 exons) Duplication 2 *XIAP* *STAG2*	Female (mother): language/learning disabilities and facial hypotonia Male sibs presenting with: severe MR language delay minor facial dysmorphism behavioural problems	(Bonnet et al. 2009)
Xq25Dup	*GRIA3* (partial – last 4 exons) *THOC2* *XIAP* *STAG2*	Male patient presenting with: severe MR skeletal (skull) abnormalities minor brain malformations distinct facial dysmorphism genitourinary and motor abnormalities behavioural problems absence of speech seizures	This study

These results suggest that additional genetic factors contribute to the syndromic MR etiology in patient XMR8, possibly through the altered gene dosage of *THOC2*, *STAG2* and *XIAP* encompassed by the duplication (Table 2.5). While none of these genes have been previously implicated in XLMR, *STAG2* is a good candidate for a contributor to the MR phenotype in this individual given its putative function as a chromatin remodeler. Duplications in other chromatin remodelers such as *ATRX* and proteins which interact with these complexes, including *MECP2*, have been

previously reported in syndromic cases of severe MR. (Lugtenberg et al. 2009, Thienpont et al. 2007, Van Esch et al. 2005). Also, by comparison to the clinical picture of these previously reported patients, patient XMR8 has a far more severe presentation suggesting a role for *THOC2* in the pathogenesis of the syndromic MR in the patient described in this study. These results suggest that *THOC2* may be a candidate MR gene, or at least play a role in embryological development.

Table 2.6 The predicted function of the genes encompassed by the Xq25Dup identified in sporadic syndromic MR patient XMR8

GENE	DISEASE-ASSOCIATION	FUNCTION
GRIA3 (OMIM ***305915**)	NS-XLMR	Glutamate receptor: AMPA-sensitive excitatory neurotransmitter receptor involved in a variety of neurophysiological functions
STAG2 (OMIM: ***604359**)	None known	Component of the cohesion complex required for sister chromatid pairing after DNA replication (*ATRX* has been proposed to have a similar function)
THOC2 (OMIM: ***300395**)	None known	THO2 forms part of the TREX (transcription/export) complex which associates with RNA polymerase II during elongation and assists mRNA export to the ribosome
XIAP (OMIM: ***300079**)	X-linked lymphoproliferative syndrome (XLP)*	Member of a protein family involved in the inhibition of apoptosis through binding of TRAF1/2 (tumour necrosis factor receptor-associated factor)

*XLP: XLP is a rare immunodeficiency characterized by extreme susceptibility to infection with Epstein-Barr virus (EBV). Symptoms include severe or fatal mononucleosis, acquired hypogammaglobulinemia, pancytopenia and malignant lymphoma

Bonnet and colleagues (2009) suggest that the second duplication (encompassing *XIAP* and *STAG2*) identified in their study does not contribute to the disease etiology due to the CNV's overlap with a known CNP (Bonnet et al. 2009). However, basing this hypothesis on the presence of a CNP should carry a note of caution. The known CNPs in the region include a CNV of unknown direction (i.e. loss/gain) which is particularly unreliable (Redon et al. 2006). The second identified CNP is a small deletion within *STAG2* (McCarroll et al. 2006). In contrast to the duplications described here, this deletion may have different pathogenic mechanisms to the duplications described in these patients (Figure 2.7). Lastly, as these CNPs were archived within the DGV database, the gender of the individuals in whom these variants were detected is unknown which may further influence the pathogenic effect of a CNV in hemizygous male patients.

Finally, the observation of three different but overlapping duplications in this Xq25 region suggests that this interval may be predisposed to structural rearrangements and may constitute a novel 'duplication hotspot'. To this end, Chiyonobu and colleagues assessed 23 male MR patients for the presence of the Xq25 duplication they described but found no further rearrangements (Chiyonobu et al. 2007). However, the small size of this cohort was a limiting factor and a larger-scale investigation in additional severely affected males with a similar phenotype to those described here will give a better indication as to the likelihood of the existence of a duplication hotspot.

Patient XMR12, presented with MR, microcephaly, minor facial dysmorphism, seizures and classic lissencephaly. Upon whole-genome SNP-array analysis an approximately 1.6Mb deletion was detected in the 4p16.3 subtelomeric region. Heterozygous deletions in this region are known to cause Wolf-Hirschhorn syndrome (WHS) (OMIM: **#194190**), a syndromic form of MR. WHS is characterised by MR, seizures, congenital heart defects, genital/renal anomalies and a distinctive facial dysmorphism which includes: microcephaly, hypertelorism, prominent glabella, broad and/or beaked nose, short philtrum, micrognathia, downturned corners of the mouth, dysplastic ears and preauricular tags (Zollino et al. 2008). The clinical presentation of patient XMR12 shows some overlap with the features of WHS, including: MR, seizures, microcephaly and some minor facial dysmorphism (broad forehead, brachycephaly, hypertelorism and a broad nasal root and tip), but does not include lissencephaly. However, it should be noted that while this deletion does lie within the chromosomal interval deleted in WHS patients, it does not overlap with the WHS critical region (WHSCR) common to all patients described to date (Zollino et al. 2008). Secondly, the genomic architecture of this subtelomeric region makes it prone to structural variation and several CNPs have been described in the DGV in this region. However, none of these CNPs overlap completely with the variant described here and none of them overlap the *ADRAC2* gene.

It should also be noted that prior karyotype analysis showed an abnormal banding pattern on chromosome 6p, the disease-causing nature of this change could not be determined as the parents were unavailable for testing. Further molecular analysis by the whole-genome SNP array analysis described here, revealed no abnormalities on

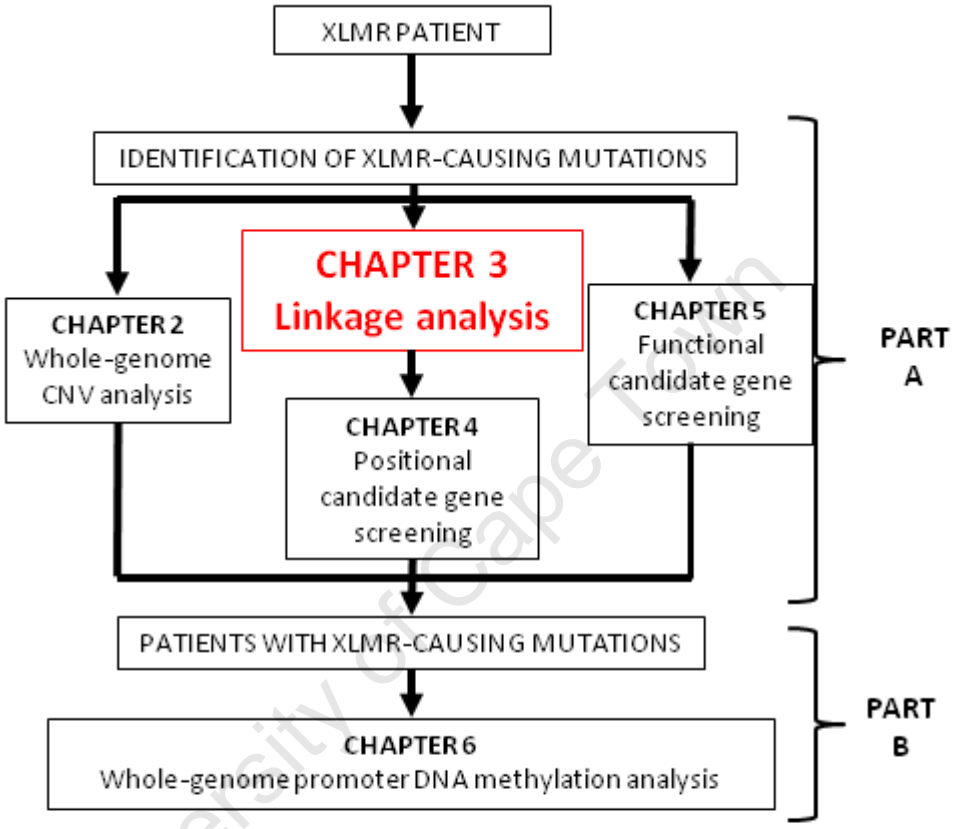
6p. The 4p16.3Del was the only other potentially disease-causing CNV detected. These results suggest that patient XMR12 may carry a 4:6 translocation (with the resultant small 4p16.3 deletion occurring at the translocation breakpoint) which may be associated with the disorder. FISH analysis using probes from either chromosome four and/or six should be performed to confirm this hypothesis.

Ideally the patient's parents need to be tested for both the abnormality on 6p and the deletion at 4p16.3 in order to elucidate the disease-causing nature of these variations. Until these analyses are done, no conclusions can be drawn as to the disease-association of the 4p16.3Del or the 6p chromosomal abnormality, nor can the parents be offered any valuable recurrence risk estimates.

In conclusion it has been illustrated that the application of the 250K Affymetrix SNP array for copy number analysis efficient in the detection of MR-causing CNVs. The exclusion criteria adopted here are, in the majority of cases (86%), effective in excluding benign CNVs from further analysis. However, the importance of using a second technique to verify variants detected by whole-genome array investigations has also been highlighted. The detection of a 9Mb X chromosomal deletion (Xq26.3-27.3Del) which exhibits low penetrance raises important questions for the genetic counselling approaches in families with known and putative disease-causing mutations. In this study we detected three disease-causing CNVs (17p11.2Del, Xq25Dup and Xq26.3-27.3Del) in the 30 MR patients assessed thus demonstrating a detection rate of 10%. This finding highlights the significant contribution that CNVs make to the MR etiology in patients with unexplained MR and emphasises the importance of implementing this technology in the MR diagnostic setting in South Africa. Moreover, the disease-causing mutations identified in these CNV analyses can now be prioritised for epigenomic profiling in Part B of this dissertation.

CHAPTER 3

LINKAGE ANALYSIS



3.1 INTRODUCTION

Linkage analysis has long been established as one of the most important tools in the identification of disease genes. In the late 1980s linkage analysis used in conjunction with information from chromosomal rearrangements led to the identification of disease genes for, amongst others, chronic granulomatous disease and Duchenne muscular dystrophy. Subsequently, using only linkage analysis and the positional candidate gene approach the gene responsible for cystic fibrosis was identified. Since then there has been a flurry of genes discovered by linkage analysis with over 40 genes indentified preceding 1995 and hundreds more subsequently (Collins. 1995). Today, this positional candidate approach is still used for the mapping of rare, autosomal recessive and complex disorders with progress being aided by the development of high-throughput whole-genome technologies as well as the wealth of genetic information generated by the human genome project.

The use of X-chromosomal linkage analysis and subsequent positional candidate gene selection to identify disease-causing mutations in four South African XLMR families was one of the primary strategies adopted in this study. The rationale for the use of this well-established technique includes both technical and practical motivations. Firstly, discovery of a fair proportion of the 87 XLMR genes identified to date was due to the successful implementation of linkage analysis in multiple XLMR families (Chiurazzi et al. 2008). This is due to the more easily identifiable X-linked inheritance pattern in MR families as well as the relative ease with which molecular techniques can be applied to the male hemizygous X-chromosomal state. Secondly, given the clinical and genetic heterogeneity associated with the disorder it is a tedious and difficult task to select an XLMR candidate gene based on a familial phenotype. Performing linkage analysis on only a single chromosome allows the definition of a critical interval within a family and can significantly reduce the number of genes which need to be investigated. For these reasons, those families which showed clear X-linked inheritance and were of sufficient size, were selected for linkage analyses in this study.

3.2 METHODS

3.2.1 FAMILY SELECTION

XLMR families with unexplained MR were selected for inclusion in this linkage analysis study for definition of a critical disease-associated chromosomal region. Family selection was based on the following inclusion criteria:

- A Clear X-linked recessive inheritance pattern
- Families negative for the FXS CGG repeat expansion mutation
- Families of sufficient size to conduct meaningful linkage analysis, i.e. a minimum of a mother and two affected sons.

Originally only two families were selected for linkage analysis, families XMR2 and Fx56. Family XMR2 consisted of over 50 relatives spanning four generations while Fx56 consisted of seven family members from three generations. Subsequently, an additional two smaller families (Fx36 and Fx67) were included in the study in an attempt to refine the critical interval in families XMR2 and Fx56. The smaller families, Fx36 and Fx67, consisted of carrier mothers as well as two and three affected brothers respectively (all pedigrees in Appendix 3A).

The clinical presentation of patients was not part of the inclusion criteria for the selection of families for linkage analysis. Nevertheless, affected XMR2 family members presented with a syndromic form of XLMR while patients in families Fx56, Fx67 and Fx36 manifested with NS-XLMR. A more detailed clinical review of patient phenotypes was conducted once a critical disease-associated chromosomal interval was defined and in order to prioritise positional candidate genes which were selected based on clinical presentation (Chapter 4).

3.2.2 MICROSATELLITE MARKER ANALYSIS

3.2.2.1 Whole X chromosomal screen

A previous study in this laboratory excluded the chromosomal region encompassing the *ARX* gene as being associated with MR in families XMR2 and Fx56 (Honours (Med) dissertation by G.Carvill, 2005). Consequently the chromosomal region surrounding *ARX* (Xptel – Xp21.2) was not included in linkage analysis in these families. Thirteen microsatellite markers spanning the remainder of the X chromosome (Xp21.2- Xqtel) were selected from the ABI® Prism Linkage Mapping Set V2 Panel 28 [Applied Biosystems] (Appendix 3B). These microsatellite markers were spaced at between 2 and 12Mb intervals (except for the 24Mb centromeric interval) so as to allow for a comprehensive screen of the X chromosome. The selected microsatellite markers were amplified from DNA of 12 and four family members of XMR2 and Fx56 respectively. Also, in family Fx67, total of 13 X-chromosomal microsatellite markers (Appendix 3B) were retrospectively genotyped in the four individuals comprising this family.

PCR amplification was performed according to manufacturer's instructions on an Applied Biosystems GeneAmp® PCR system 9700 thermal cycler and products were visualised with ethidium bromide (EtBr) on a 1% agarose gel with a 100bp molecular weight (MW) marker [Fermentas] (Appendix 3C) in order to establish PCR success. Genotyping of microsatellite markers was achieved on an ABI Prism® 3100 Genetic Analyser [Applied Biosystems] in accordance with manufacturer's protocols and analysed using the GeneMapper™ V3.0 [Applied Biosystems] software.

3.2.2.2 Fine mapping markers

Once a chromosomal region indicative of an association with the disorder was established in XMR2, an additional ten microsatellite markers within the applicable interval (between DXS993 and DXS986) were selected in order to refine this critical region (Appendix 3B). Furthermore, four family members were recruited to this study and genotyped, in order to refine this disease-associated interval. The 10 microsatellite markers were genotyped in all 16 members of family XMR2.

Fine mapping at the same critical interval (between DXS993 and DXS986) was also conducted in family FX56 using six of the aforementioned microsatellite markers (Appendix 3B). Finally, in families Fx36 and Fx67, these six fine mapping microsatellite markers, as well as the microsatellite markers within the disease-associated interval (DXS993, DXS991, DXS986) were genotyped in these families in an attempt to refine the critical interval in families XMR2 and Fx56.

Retrospectively, an additional eight fine-mapping microsatellite markers (between DXS987 and DXS993 and at a second region between DXS990 and DXS1106) were genotyped in family Fx67 in an attempt to refine the chromosomal regions suggestive of disease-association in this family (Appendix 3B).

All microsatellite markers were selected using the uniSTS database (NCBI) (http://www.ncbi.nlm.nih.gov/). Selected markers were either commercially available (Research Genetics) or alternatively, synthesised using a Beckman Oligo 1000M DNA synthesizer and labelled with a 5' fluorescent tag [synthesised by P. Ma, UCT]. A range of PCR protocols were used to amplify the selected microsatellite markers (Appendix 3D) on a Px2 thermal cycler [Thermo Electron Corporation]. Gel electrophoresis and genotyping analysis was performed as above.

3.2.3 TWO-POINT LINKAGE ANALYSIS

Two-point linkage analyses were performed for each microsatellite marker using the MLINK programme from the LINKAGE software package (Version 5.1.). Full penetrance of the XLMR gene was assumed and the frequency of the disease allele was set at 0.0001, otherwise all analyses were conducted as previously described (Terwilliger and Ott. 1994).

3.3 RESULTS

The results of the two-point linkage analyses for the 13 microsatellite markers used in the initial whole X chromosomal genotyping performed on DNA samples from 12 individuals comprising family XMR2 is shown in Table 3.1. These preliminary findings suggested linkage to the chromosomal region encompassing microsatellite marker DXS991 with a maximum LOD score of 2.83 (θ = zero) at DXS991 (Table 3.1). This chromosomal region (Xp11.4-Xq21.1) spanned an approximately 38Mb chromosomal region between (but excluding) microsatellite markers DXS993 and DXS986

Table 3.1 Two-point LOD scores for microsatellite markers from initial X-chromosomal screen in Family XMR2*

MARKER	LOD SCORES AT APPLICABLE θ VALUES					
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
DXS1214	-∞	-0.53	-0.01	0.13	0.11	0
DXS1068	-∞	-0.53	-0.01	0.13	0.11	0
DXS993	-∞	0.62	0.58	0.39	0.15	0
DXS991	**2.83**	**2.34**	**1.79**	**1.19**	**0.56**	**0**
DXS986	-∞	-2.29	-1.18	-0.6	-0.23	0
DXS990	-∞	1.38	1.20	0.86	0.46	0
DXS1106	-∞	0.43	0.60	0.50	0.28	0
DXS8055	uninformative					
DXS1001	-∞	-0.53	-0.01	0.13	0.11	0
DXS1047	-∞	-0.53	-0.01	0.13	0.11	0
DXS1047	-∞	-0.53	-0.01	0.13	0.11	0
DXS1227	-∞	-0.78	-0.21	-0.02	0.03	0
DXS8043	-∞	-0.07	0.12	0.15	0.11	0
DXS8091	-∞	-1.65	-0.67	-0.22	-0.02	0

*maximum LOD score for 12 individuals (grandmother's genotype inferred) in this family was 2.83

Linked microsatellite markers appear in bold

Haplotypes were constructed from the genotyping data of the 13 microsatellite markers used in this initial whole X-chromosomal screen (Figure 3.1).

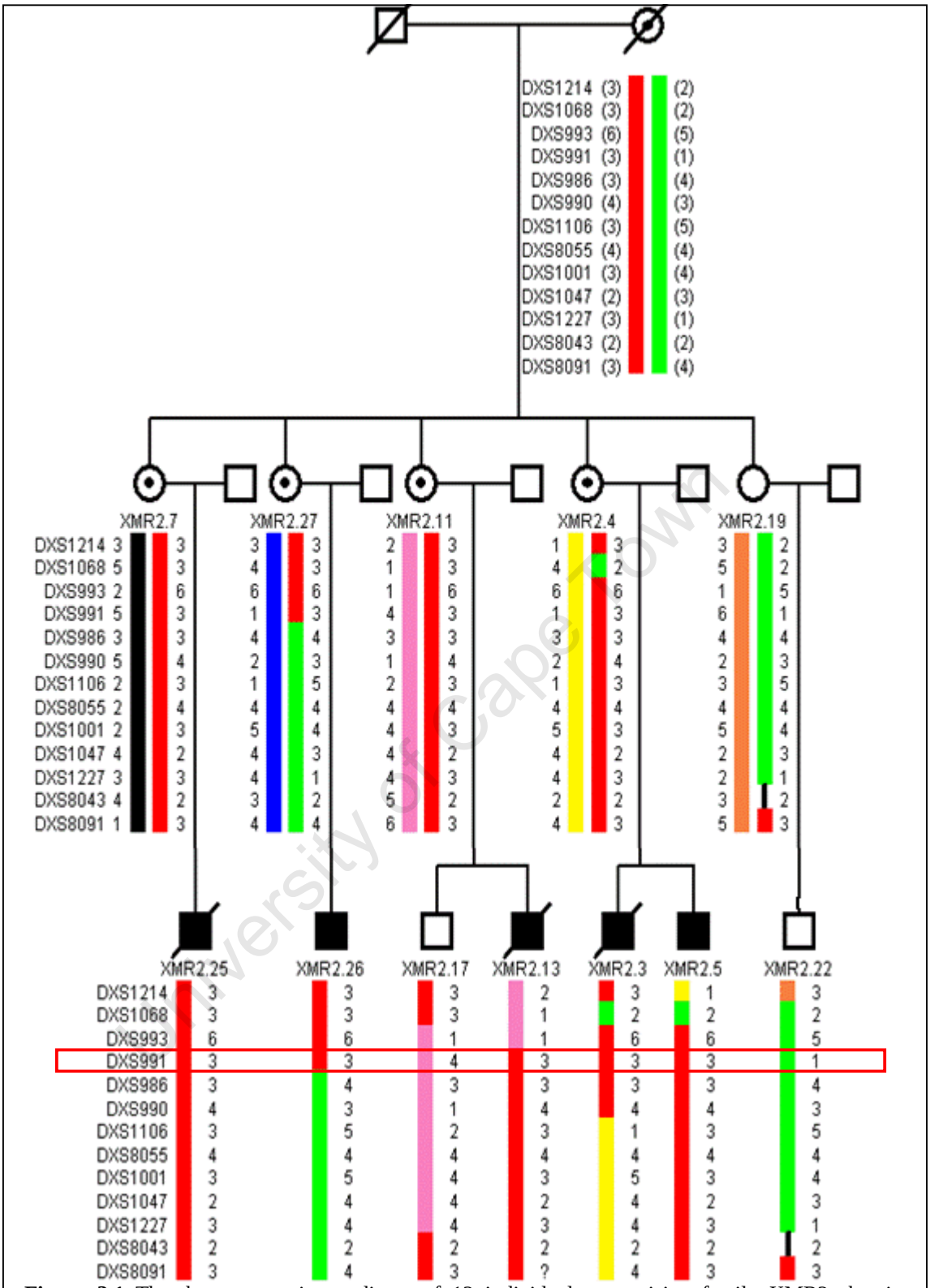


Figure 3.1 The three generation pedigree of 12 individuals comprising family XMR2 showing individual haplotypes from the initial whole X chromosomal analysis. The maternal grandmother's genotype was inferred as designated by the allele occurring in brackets e.g. (1). Haplotypes are represented by solid coloured bars, with the disease-associated haplotype represented in red. The red rectangle encompasses the region of positive linkage at microsatellite marker DXS991 with a maximum LOD score of 2.83. It should also be noted that there was a significant degree of non-paternity in this family, with each of the five sisters (and the sixth in Figure 3.2) having a different father. As the XLMR in this family was due to an X-chromosomal mutation, which segregates through the maternal line, the non-paternity in this family did not affect the linkage analyses.

Prior to fine mapping of the 38Mb disease-associated interval, an additional three family members were recruited for linkage analysis in this family (XMR2.8/29/30/31). The results of the two-point linkage analyses using the ten fine-mapping microsatellite markers in these 16 family members of XMR2 is given in Table 3.2. The critical interval was refined to a 10Mb region between (but not including) microsatellite markers DXS8111 and DXS986 (Xq13.1-q21.1) with a maximum LOD score of 3.08 (θ = zero) across the region (i.e. DXS8052-DXS56).

Table 3.2 Two-point LOD scores in family XMR2 for fine mapping microsatellite markers

MARKER	**LOD SCORES AT APPLICABLE θ VALUES**					
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
DXS6810	-∞	0.32	0.42	0.36	0.22	0
G10578	-∞	0.99	0.89	0.63	0.29	0
DXS1199	1.03	0.87	0.69	0.49	0.27	0
DXS1190	1.33	1.14	0.92	0.67	0.37	0
DXS6785	2.36	1.95	1.49	0.99	0.46	0
DXS8111	-∞	1.84	1.55	1.09	0.54	0
DXS8052	**3.08**	**2.50**	**1.89**	**1.24**	**0.58**	**0**
DXS559	**3.08**	**2.50**	**1.89**	**1.24**	**0.58**	**0**
DXS441	**3.08**	**2.52**	**1.92**	**1.27**	**0.61**	**0**
DXS56	**3.08**	**2.52**	**1.92**	**1.27**	**0.61**	**0**

*maximum LOD score for 13 individuals (grandmother's genotype inferred) in this family was 3.08
Linked microsatellite markers with the highest LOD score appear in bold

Haplotypes for these 16 individuals were generated from the genotypes of the ten microsatellite markers, the haplotype regions encompassing the critical interval are shown in Figure 3.2. Interestingly, the unaffected individual, XMR2.31, inherited the same haplotype as the affected relatives until a recombination event between microsatellite markers DXS8111 and DXS8052 (Figure 3.2). This recombination ultimately resulted in the reduction of the linked interval to the 10Mb chromosomal region at Xq13.1-q21.1 (between microsatellite markers DXS8111 and DXS986).

No additional family members were available for genotyping analysis making it impossible to further refine this 10Mb critical interval, therefore this region was interrogated for positional candidate genes (Chapter 4).

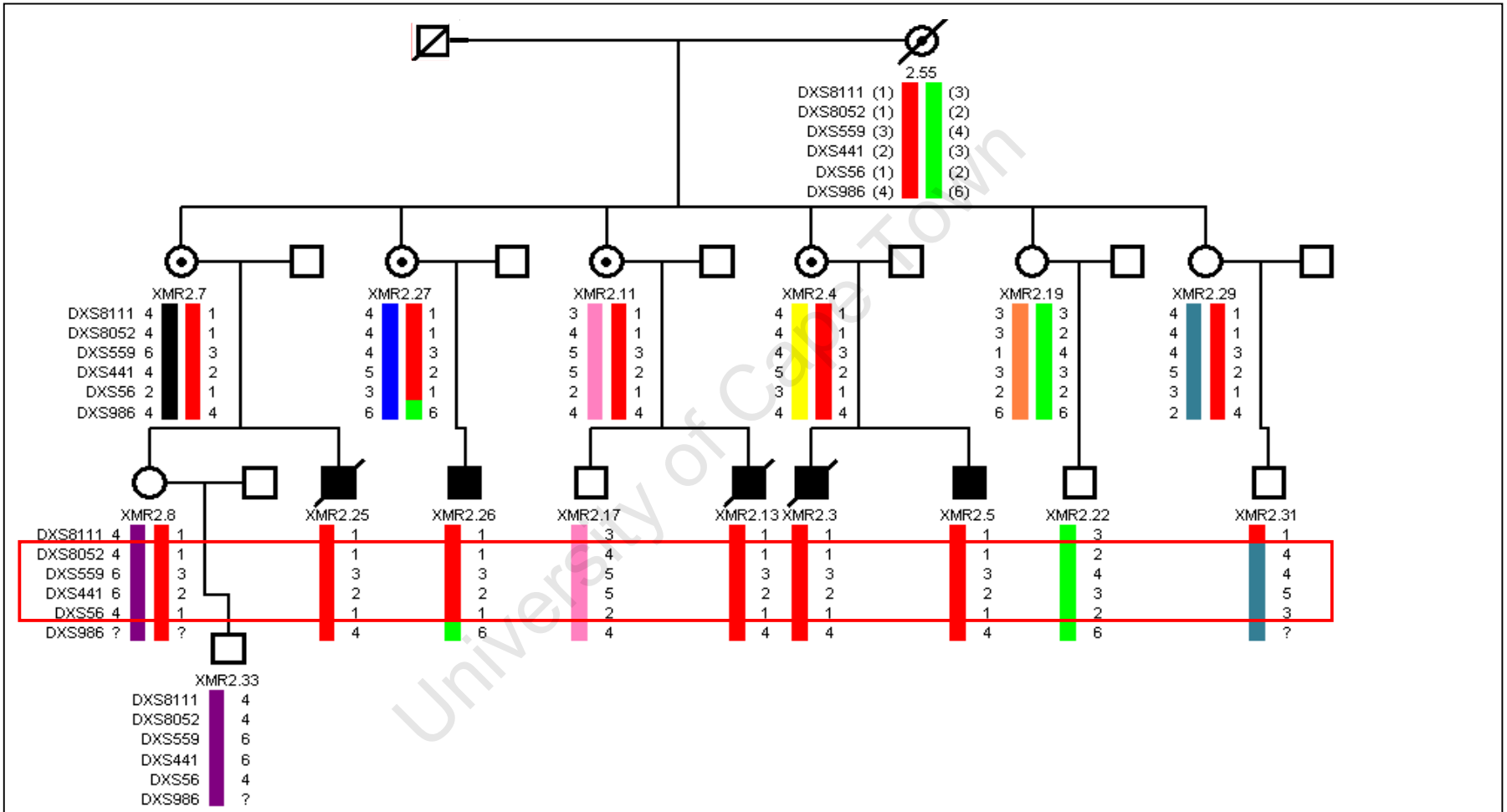


Figure 3.2 The four generation pedigree of selected individuals comprising family XMR2 showing fine-mapping of the critical interval. The maternal grandmother's genotype was inferred as designated by the allele in brackets e.g. (1). Haplotypes are represented by solid coloured bars, the disease-associated haplotype is represented in red. The red rectangle encompasses the region of positive disease-association between microsatellite markers DXS8111 and DXS986 with a LOD score of 3.08. It should also be noted that there was a significant degree of non-paternity in this family, with each of the six sisters having a different father.

The whole X-chromosomal linkage analysis in 13 microsatellite markers in family Fx56 led to the identification of a disease-associated chromosomal region that overlapped with family XMR2. Fine mapping of an additional six microsatellite markers within this region in family Fx56 ultimately led to the identification of a disease-associated interval between (but not including) microsatellite markers DXS1068 and DXS8055. The maximum LOD score of 0.60 (θ=zero) was achieved for the chromosomal interval encompassing markers DXS993-DXS986 (Table 3.3). Although this LOD score was not statistically significant, this chromosomal region could not be excluded due to the co-segregation of alleles for markers DXS1068-DXS8055 (and DXS1227) in affected individuals in this three generational pedigree.

Table 3.3 Two-point LOD scores in family Fx56 for X-chromosomal microsatellite markers

MARKER	LOD SCORES AT APPLICABLE θ VALUES					
	0	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
DXS1214	-∞	-0.70	-0.40	-0.22	-0.10	0
DXS1068	-∞	-0.48	-0.26	-0.16	-0.08	0
DXS993	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS6810	**uninformative**					
G10578	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS1199	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS991	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS1190	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS6785	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS559	**uninformative**					
DXS986	**0.60**	**0.47**	**0.34**	**0.21**	**0.1**	**0**
DXS990	**0.30**	**0.26**	**0.20**	**0.15**	**0.08**	**0**
DXS1106	**0.30**	**0.21**	**0.13**	**0.06**	**0.02**	**0**
DXS8055	-∞	-0.19	0.01	0.07	0.06	0
DXS1001	-∞	-0.44	-0.19	-0.08	0.02	0
DXS1047	-∞	-0.19	0.01	0.07	0.06	0
DXS1227	0.30	0.26	0.20	0.15	0.08	0
DXS8043	-∞	-0.19	0.01	0.07	0.06	0
DXS8091	-∞	-0.19	0.01	0.07	0.06	0

*maximum LOD score for 4 individuals in this family was 0.60
Linked microsatellite markers with the highest LOD score appear in bold

Haplotypes were constructed from the genotypes of the 19 X-chromosomal microsatellite markers (initial whole-X screen and fine mapping markers) in family Fx56 (Figure 3.3). Unfortunately, no further family members were available for genotyping analysis and thus the disease-associated interval could not be reduced in this family, nor the LOD score increased. The disease-associated chromosomal region (Xp11.4-Xq21.1) spanned approximately 76Mb and contains roughly 397 genes (based on March 2006 assembly) of which 33 have been implicated in XLMR.

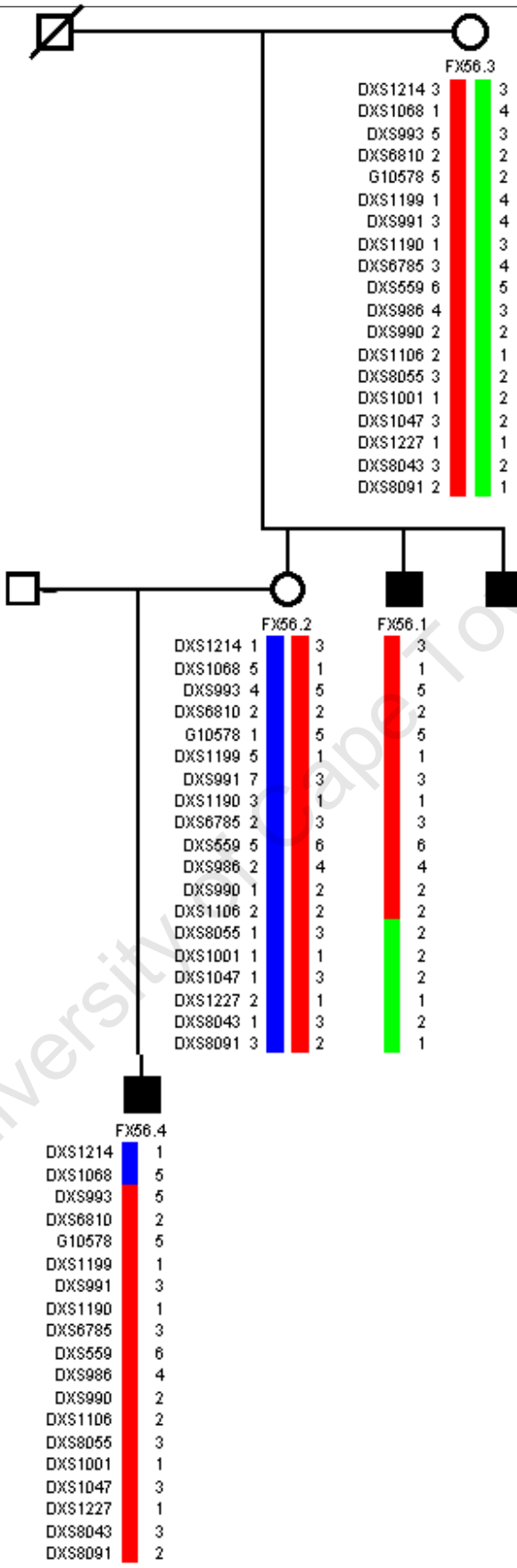


Figure 3.4 The three generation pedigree of Fx56, showing haplotype analysis from the genotyping of 19 X-chromosomal microsatellite markers. Haplotypes are represented by solid coloured bars, with the disease-associated haplotype represented in red. The disease-associated region was defined by the recombination between markers DXS1068 and DXS993 in individual FX56.4 on the Xp arm and between markers DXS1106 and DXS8055 in individual FX56.1 on the Xq arm.

Genotyping analysis using nine microsatellite markers spanning Xp11.4-Xq21.1 was originally conducted in the smaller families, Fx36 and Fx67, in order to refine the critical interval originally identified in families XMR2 and Fx56. The combined data (from all four families) was assessed in an attempt to refine the critical interval in families XMR2 and Fx56. Haplotype analysis in family Fx36 showed allele co-segregation across the entire investigated chromosomal interval, Xp11.4-Xq21.1 (Figure 3.4). Conversely, in family Fx67 the region Xp11.4-Xq21.1 was shown not to segregate with the disorder (Figure 3.5). Taken together, these results were not successful in refining the critical interval in families XMR2 and Fx56.

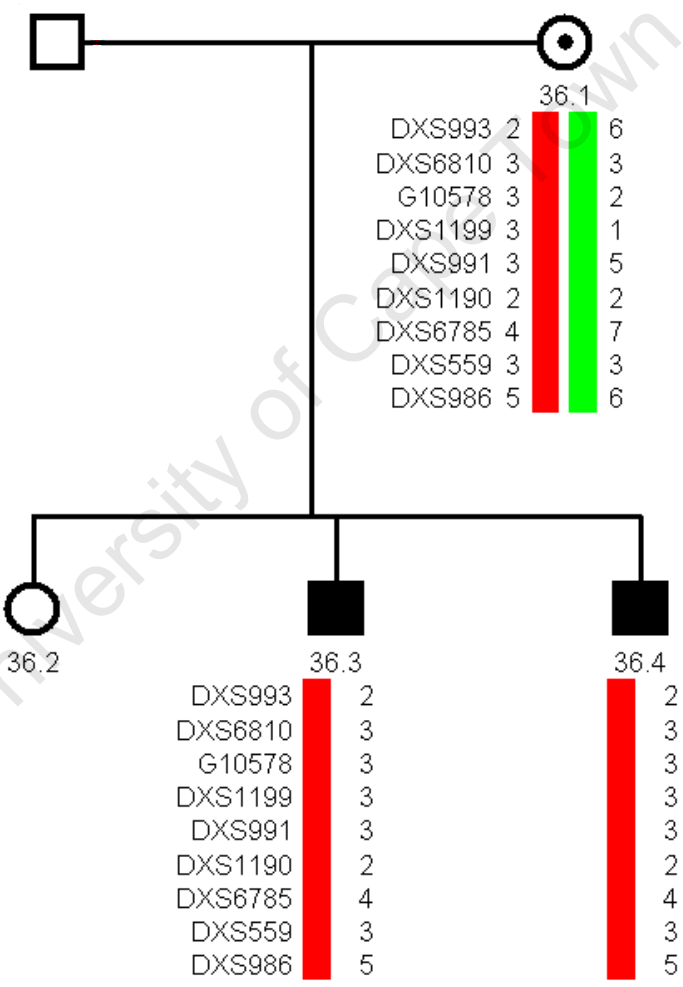


Figure 3.4 The two generation pedigree of selected individuals illustrating the allele segregation in family Fx36 for the nine microsatellite markers genotyped in the Xp11.4-Xq21.1 region. Haplotypes are represented by solid coloured bars, with the disease-associated haplotype represented in red. In this study the two both affected brothers (Fx36.3 and 36.4) were shown to have inherited the same haplotype for this particular chromosomal region; therefore this interval (Xp11.4-Xq21.1) may harbour the disease-causing mutation in family Fx36

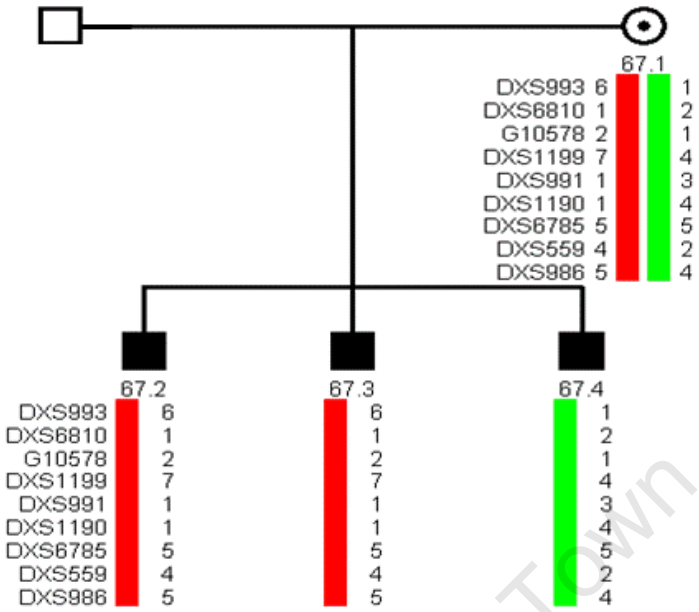


Figure 3.5 The two generation pedigree of selected individuals illustrating the allele segregation in family Fx67 for the nine microsatellite markers in the Xp11.4-Xq21.1 region. Haplotypes are represented by solid coloured bars. In this study two of the affected brothers (Fx67.2/3) inherited the same haplotype (red) for this particular chromosomal region. However, in contrast, affected individual Fx67.4 has inherited a different haplotype (green), therefore, this allele segregation analysis demonstrated that this chromosomal region was unlikely to harbour the disease-causing mutation.

No indication of disease-association in family Fx67 was evident at the nine microsatellite markers spanning Xp11.4-Xq21.1. However, genotyping of an additional ten X-chromosomal microsatellite markers in family Fx67 demonstrated that only a single marker, DXS1068, segregated with the disorder in the three affected brothers. A significant LOD score cannot be generated due to the small size of the family (the maximum LOD score for this family is 0.40 (θ=zero)). However, co-segregation of DXS1068 with the disorder suggested this region was associated with MR in Fx67. Therefore, subsequent genotyping of seven microsatellite markers flanking marker DXS1068, demonstrated that a total of five microsatellite markers segregated with the disorder in this family (though genotyping of one of these markers, DXS1214, was uninformative in this family) (Figure 3.6). This putatively disease-associated region spanned the interval between (but not including) markers DXS9896 and DXS990. This chromosomal interval spans approximately 12Mb and is relatively gene poor; therefore, this region was further assessed for positional candidate genes in Chapter 4.

It should also be noted that while DXS1068 was the only microsatellite marker which segregated with the disorder in this family, there was an additional chromosomal interval which could not be excluded due to a recombination between markers DXS990 and DXS1106. An additional marker (DXS6789) within this region was genotyped and led to the identification of another putative disease-associated region between but not including DXS990 and DXS6789 (Figure 3.6). This region could not be further refined, however, this chromosomal interval contains no known genes and was therefore not investigated further.

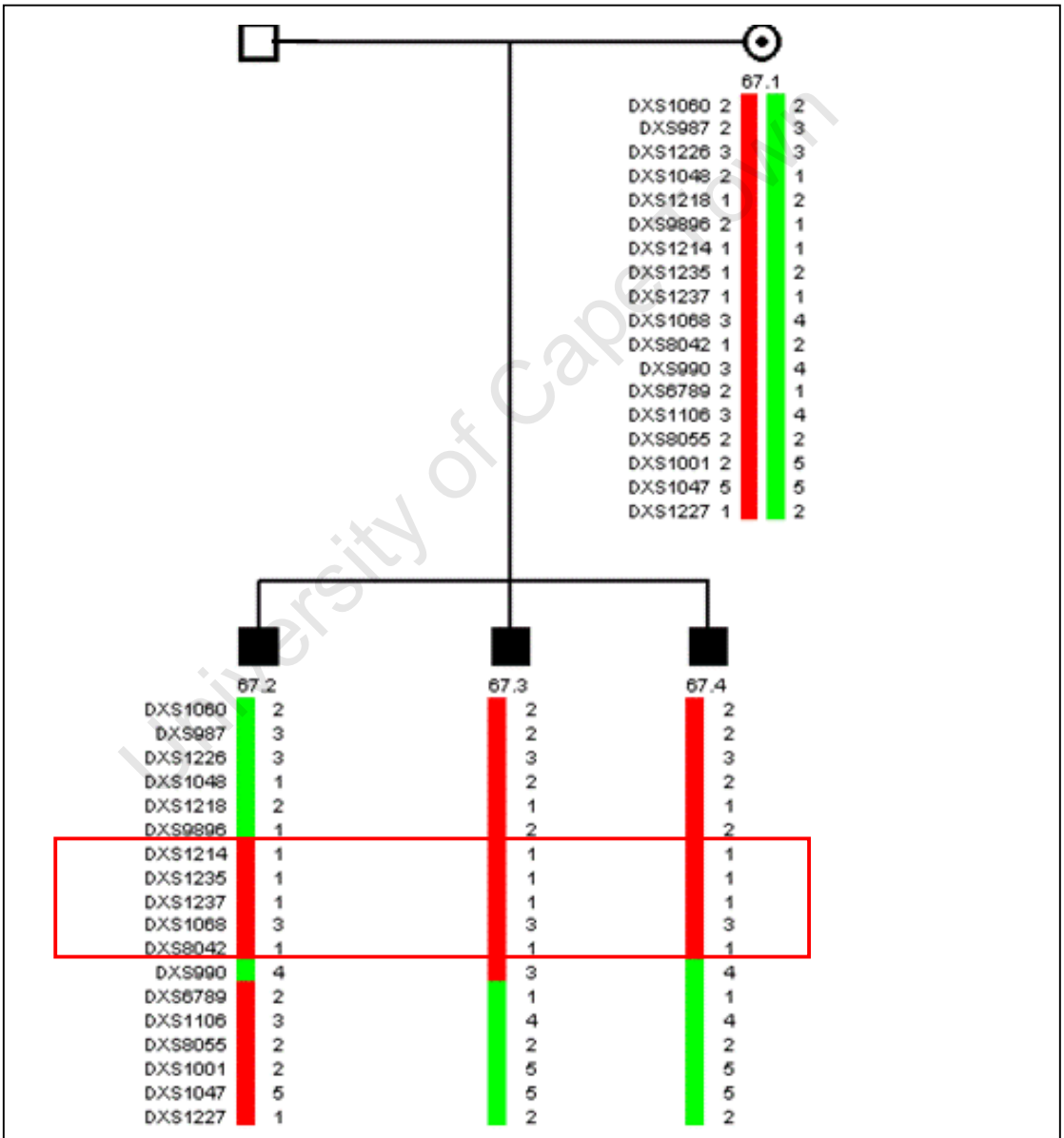


Figure 3.6 The two generation pedigree of selected individuals illustrating the allele segregation in family Fx67 for the 18 microsatellite markers spanning the remainder of the X chromosome (excluding Xp11.4-Xq21.1). Haplotypes are represented by solid coloured bars, the 'affected' allele is represented in red whereas the normal allele is represented in green. In this study, five microsatellite markers were shown to segregate with the disorder (red rectangle) in family Fx67. These results suggested that this region may harbour the disease-causing mutation.

3.4 DISCUSSION

To date, linkage analysis has been successfully applied in the identification of a large proportion of the 87 known XLMR genes (Chiurazzi et al. 2008). In this study, the linkage analysis technique was employed in order to identify a disease-associated X-chromosomal interval in four South African XLMR families, with a view towards positional candidate gene selection and subsequent mutation detection.

In the largest of the investigated families, XMR2, initial whole X-chromosomal microsatellite analysis in 12 individuals led to the identification of a 38Mb region with a maximum LOD score of 2.83 (θ=zero), at marker DXS991. This LOD score was statistically significant (i.e. LOD score <2, for the X-chromosome), and was thus indicative of disease-association of this locus in family XMR2. Therefore, fine mapping in this region was conducted including four additional family members (total of 16 individuals) in order to refine the disease-associated interval. A recombination event in a newly-recruited family member (XMR2.31) reduced the interval to a chromosomal region spanning approximately 10Mb with a maximum LOD score of 3.08 (θ=zero) across the linked interval (i.e. DXS8052-DXS56). The increase in LOD score reflects the inclusion of additional family members in the analysis, as well as the recombination in individual XMR2.13. This statistic was the maximum LOD score that could be achieved for a family of this size. The fine mapping in family Fx67 reduced the critical interval to a 10Mb chromosomal region (Xq13.1-q21.1), flanked by (but not including) the microsatellite markers DXS8111 and DXS986. Given that the LOD score of 3.08 (θ=zero) was also statistically significant, it was highly probable that this chromosomal region harboured the disease-causing mutation in this family. The selected positional candidate genes within this disease-associated interval are discussed in further detail in Chapter 4.

In family Fx56, the initial whole X-chromosomal screen identified a disease-associated interval (Xp11.4-Xq21.1) between (but excluding) the microsatellite markers DXS1068 and DXS8055. The LOD score for this chromosomal region in family Fx56 was 0.60 (θ = zero), achieved for the chromosomal interval encompassing markers DXS993-DXS986. This statistic was also the maximum LOD score which can be achieved for a family of this size. Although this LOD score

was not statistically significant, this was a reflection of the small family size rather than any genetic factor. Unfortunately, additional Fx56 family members could not be recruited and therefore further fine mapping of the disease-'associated' X-chromosomal interval could not be conducted. This disease-associated chromosomal interval spans approximately 76Mb, encompassing roughly 397 genes of which 33 have been previously implicated in the pathogenesis of XLMR. Of these 33 XLMR genes, 11 (*PAK3*, *ACSL4*, *NLGN3*, *DLG3*, *ARHGEF9*, *JARID1C*, *FTSJ1*, *ZNF81*, *ZNF41*, *ZNF674*, *EFHC2*) have been shown to be mutated in patients with NS-XLMR. Given the non-syndromic clinical presentation of affected individuals in family Fx56, it was difficult to prioritise NS-XLMR genes for mutation detection analysis in the absence of additional clinical features. For these reasons it was deemed neither financially nor technically viable at this point to conduct positional candidate gene analysis in this particular XLMR family.

The X-chromosomal allele segregation analysis in families Fx36 and Fx67 was originally conducted in an attempt to refine the linkage interval identified in families XMR2 and Fx56 (i.e. Xp11.4-Xq21.1) using combined data. Haplotype analysis across the Xp11.4-Xq21.1 interval in family Fx36 showed this entire chromosomal region segregated with the disorder in this family. Given that all nine microsatellite markers (DXS993-DXS986) segregated with the disorder in this family, this information was not informative with respect to refining the critical interval in families Fx56 and XMR2. Conversely, in family Fx67, haplotype analysis in the same region (Xp11.4-Xq21.1) showed no disease-association with this chromosomal region. Due to the lack of recombination events in families Fx36 and Fx67 within Xp11.4-Xq21.1, haplotype analysis of this region was ultimately ineffective in narrowing the linkage interval in families XMR2 and Fx56.

One could question the validity of using information from multiple families to refine a critical region for gene identification in XLMR families. While this is common practice for clinically homogenous disorders it is particularly difficult to apply to XLMR given the high degree of genetic and clinical heterogeneity. This is especially true in this case where these XLMR families have a diverse and variable clinical presentation and are unlikely to have the same underlying pathogenic mutation. However, in defence of the adopted strategy it has been shown that several

XLMR genes (e.g. *ARX*, *SLC6A8*, *MECP2*, *OPHN1*, *ATRX*) are implicated in the pathogenesis of both syndromic and non-syndromic forms of the disorder (Chiurazzi, Tabolacci and Neri. 2004).

Family Fx67 was not linked to the aforementioned Xp11.4-Xq21.1 chromosomal region investigated in this study using nine microsatellite markers. Further genotyping and haplotype analysis of microsatellite markers spanning the remainder of the X chromosome showed the marker DXS1068 segregated with the disorder in this family. Despite the fact that a statistically significant LOD score could not be obtained for due to the small family size (the maximum LOD score for this family is 0.40 (θ=zero), allele segregation analysis demonstrated that a total of five microsatellite markers segregated with the disorder (DXS1214-DXS8042). This 12Mb interval (Xp21.1-Xp11.4) is relatively gene poor with 27 known genes (based on human genome assembly 36, March 2006), of which only two have been previously implicated in XLMR. These two XLMR genes, *ATP6AP2* and *TM4SF2* have been shown to be mutated in NS-XLMR making them good positional candidate genes given the non-syndromic presentation of MR in family Fx67. Therefore, these two XLMR genes were assessed for DNA sequence variations as described in Chapter 4.

One of the major caveats in these linkage analyses was the limited number of family members available for genotyping analysis. The only family of sufficient size to produce a statistically significant LOD score, was family XMR2 which included genotype information from 16 individuals. For the remaining families the LOD scores were all below one, resulting in a less than 10 times greater chance that the observed linkage is real as opposed to occurring randomly. Genotyping analysis in an increased number of individuals would have a directly proportional effect on the LOD score. However, all three families with low LOD scores, FX56/36/67, could not be contacted or visited and therefore additional family members could not be recruited. In future studies it would be recommended that frequent contact is maintained with families prior to embarking on whole X-chromosome linkage analysis.

As previously mentioned, linkage analysis is a well established technique, it could be argued that its use here is outdated and novel high-throughput methods would be better suited for XLMR gene identification. It is true that the application of high-throughput sequencing and microarray technologies holds much promise for the diagnosis of MR. The speed with which many different genes can be analysed is a huge advantage particularly in the case of XLMR where diagnosis is complicated by high degrees of clinical and genetic heterogeneity. This is particularly true in NS-XLMR cases where a lack of additional clinical features makes the selection of one or several candidate genes from a host of NS-XLMR genes virtually impossible.

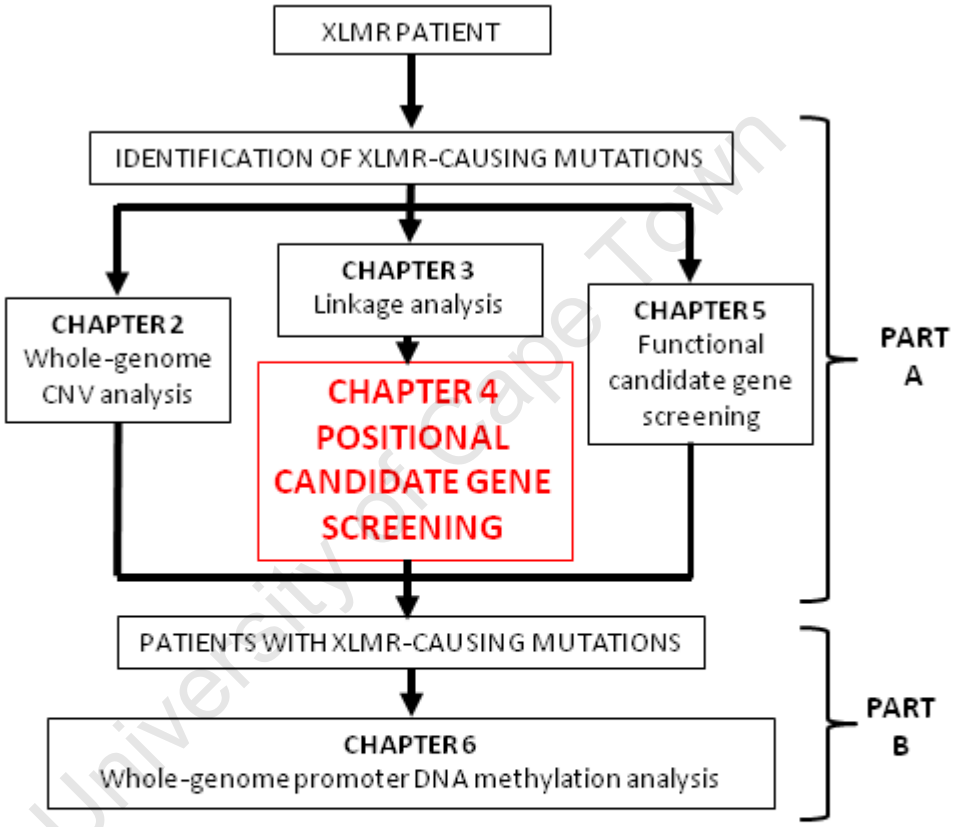
Recently, the advantages of these technologies have been illustrated in an investigation of 208 XLMR families by high-throughput Sanger sequencing which led to the discovery of nine novel of disease-associated genes (Tarpey et al. 2009). While several of the South African XLMR families described in this dissertation were included in this study, the cost of these technologies is prohibitive for diagnostic applications in a developing country such as South Africa. It is envisaged that as the costs of these technologies are reduced in the future, it will become possible to conduct mutation detection in all XLMR genes in patients as a routine diagnostic test, both in South Africa and the larger international community.

Nevertheless, this ideal diagnostic situation will not be a reality for a number of years and alternative forms of molecular diagnoses of MR need to be expanded in the South African setting. While linkage analysis cannot always lead to the identification of a disease-causing mutation, it can give an indication of the chromosomal region most likely associated with the disease phenotype in a particular XLMR family. The haplotype analyses described in this study could be used for carrier testing for individuals in family XMR2 in the absence of a known familial disease-causing mutation. This information is vital in better genetic management and counselling for families with a history of XLMR.

In conclusion, two independent critical intervals were identified in families Fx67 and XMR2 and positional candidate genes could be selected. Here it has been demonstrated that the linkage analysis technique is still a valuable way to eliminate XLMR genes and hence reduce the number of candidate genes.

CHAPTER 4

POSITIONAL CANDIDATE GENE MUTATION DETECTION



4.1 INTRODUCTION

Subsequent to the definition of a critical chromosomal interval by linkage analysis, positional candidate gene selection is necessary to identify the molecular basis of the underlying familial MR etiology. In this study, positional candidate genes were selected from the linked chromosomal interval in family XMR2 and the disease-associated region in Fx67. These preliminary linkage analyses significantly reduced the number of XLMR-candidate genes in each family from the 87 known XLMR genes. Positional candidate genes were not selected in families Fx56 and Fx36 as the disease-associated chromosomal intervals in these families spanned large regions and did not sufficiently reduce the number of XLMR-candidate genes.

4.1.1 POSITIONAL CANDIDATE GENE SELECTION: FAMILY XMR2

Linkage analysis in family XMR2 identified a 10Mb critical interval located at Xq13.1-q21.1 with a maximum LOD score of 3.08 (θ=zero) across the region (i.e. DXS8052-DXS56) (Chapter 3). This critical interval spanned the chromosomal region flanked by markers DXS8111 and DXS986. The linked region encompassed ~70 genes, nine of which (*DLG3, MED12, NLNGN3, SLC16A2, KIAA2022, ATRX, ATP7A, IAP* and *PGK1*) have been previously associated with XLMR. These nine XLMR genes were prioritised for successive mutation screening in affected individuals of family XMR2.

A number of the XLMR genes in the Xq13.1-q21.1 chromosomal interval are associated with clinical features which did not overlap considerably with the phenotype of affected individuals in family XMR2. On this basis, a number of candidate XLMR genes were not deemed a high priority for mutation screening, including: *DLG3* (NS-XLMR) (OMIM **+300189**), *MED12* (FG syndrome) (OMIM ***300188**), *NLNGN3* (X-linked Asperger syndrome) (OMIM ***300336**), *ATP7A* (X-linked Menkes disease) (OMIM ***300011**) and *PGK1* (phosphoglycerate kinase-1 deficiency) (OMIM ***311800**). Also, mutations in *SLC16A2* are known to cause Allan-Herndon-Dudley syndrome (AHDS) (OMIM ***300095**). While AHDS features showed some phenotypic overlap with the clinical presentation of XMR2 affected males, all AHDS patients exhibit elevated serum T3 (Triiodothyronine) levels (Raymond. 2006). In contrast, in this family, an affected individual (XMR2.5) had

normal levels of this thyroid hormone upon biochemical investigation, therefore *SLC16A2* was excluded as a candidate gene for MR in this family.

Mutation detection in *KIAA2022* also did not take precedence, as the exact role of *KIAA2022* in the pathogenesis of XLMR has not been fully established. To date, only a single *KIAA2022* mutation has been described in two MR patients with an X-inversion (which also disrupts another gene, *P2RY8*) (Cantagrel et al. 2004, Cantagrel et al. 2004). Finally, mutations in *IAP* were only identified in 2008 after the conclusion of the positional candidate gene mutation screening in family XMR2; IAP was therefore never considered a candidate in this family (Molinari et al. 2008).

The only remaining XLMR candidate gene in the linked region was *ATRX* (OMIM ***300032**), mutations in which are associated with a syndromic form of XLMR. Upon thorough review of the phenotype of the affected individuals in family XMR2, there was an adequate degree of clinical overlap with known ATR-X patients to select ATRX as the first candidate gene for mutation detection in this family.

4.1.1.1 Positional candidate gene: ATRX

Phenotypes associated with *ATRX* mutations

Mutations in *ATRX* were first identified in X-Linked alpha thalassaemia mental retardation (ATR-X) syndrome. This syndrome is characterised by severe mental retardation, alpha thalassaemia, facial dysmorphism and urogenital abnormalities (Weatherall et al. 1981). However, the clinical spectrum associated with *ATRX* has since been expanded to include the following allelic conditions:

- Carpenter-Waziri syndrome (Abidi et al. 1999)
- Juberg-Marsidi syndrome (Villard et al. 1996)
- NS-XLMR without alpha thalassaemia and/or epilepsy (Guerrini et al. 2000, Villard et al. 1996, Wieland et al. 2005, Yntema et al. 2002)
- XLMR with spastic paraplegia (Lossi et al. 1999)
- Holmes-Gang syndrome (Stevenson et al. 2000)
- Smith-Fineman-Myers syndrome (Villard et al. 2000)
- Chudley-Lowry syndrome (Abidi et al. 2005).

This wide range of clinical conditions caused by *ATRX* mutations complicates a clinical diagnosis. However, the prevalence of particular clinical features in *ATRX* mutation positive patients, as provided by Gibbons (2006) may serve as a guide for conducting *ATRX* screening. These features include: profound MR (95% cases), characteristic facial features (<90% cases), genetic abnormalities (80% cases), alpha thalassaemia (90% cases), skeletal abnormalities (90% cases), microcephaly (75% cases), short stature (65% cases), seizures (30% cases), cardiac defects (20% cases) and renal/urinary abnormalities (15% cases) (Gibbons. 2006).

ATRX: The gene and its protein

Mutations in *ATRX* (also known as *XNP* or *XH2*) and the resulting S-XLMR phenotype were first identified in 1995 (Gibbons et al. 1995). The gene, located at Xq13 spans approximately 300kb of genomic DNA and contains 36 exons (Picketts et al. 1996).

The *ATRX* gene encodes for a number of alternately spliced mRNA transcripts. There are at least two larger ~10.5kb transcripts which are variable at the 5'end and give rise to 265kDa and 280kDa proteins (Picketts et al. 1996). An additional shorter transcript (~7kb) has also been described, this isoform retains intron 11, resulting in a truncated protein (Garrick et al. 2004). *ATRX* is expressed in multiple foetal and adult tissues including brain, lung, kidney, liver as well as adult lymphocytes (Villard et al. 1997).

The ATRX protein is a member of the mating type switching/sucrose non-fermenting (SWI/SNF) family of chromatin associated proteins. There are two principle domains that comprise the structure of ATRX, the N-terminus ATRX-DNMT3-DNMT3L (ADD) domain and the C-terminal helicase domain (Figure 4.1) (Aapola et al. 2000). The ADD domain is comprised of three regions: a GATA-like zinc finger, a PHD (plant homeodomain) –like zinc finger and a C-terminal α-helix domain (Argentaro et al. 2007). Collectively these motifs may act as DNA or chromatin binding sites (directly or indirectly). This hypothesis is supported by several lines of evidence:

- The domain has high sequence and structural similarity to the de novo DNA methyltransferases (DNMT3A/3B/L) which possess DNA binding abilities (Xie et al. 1999)
- PHD zinc fingers are predicted to tether proteins either directly or indirectly to chromatin (Bienz. 2006).
- The structure of the GATA-like zinc finger has DNA binding potential (Argentaro et al. 2007).

These investigations and observations would suggest a DNA and protein-protein interacting function for ATRX but to date the nature of this relationship remains elusive.

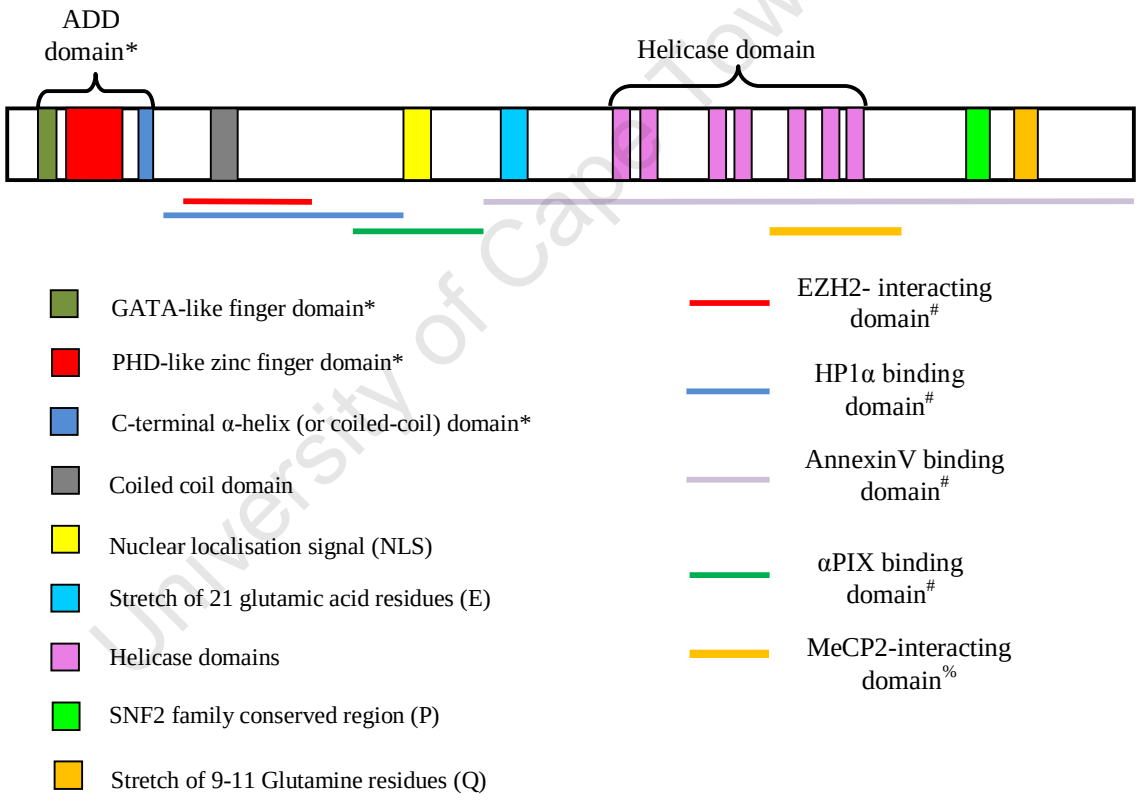


Figure 4.1 A schematic representation of the domains of the ATRX protein (not to scale). (adapted from Tang 2006, *ADD domain details from Argentaro et al. 2007, #interacting protein domain details from Park et el. 2004, %details from Nan et al. 2007.

At the C-terminus lies the helicase domain which belongs to the SNF2 (SWI/SNF) family of chromatin associated proteins (Picketts et al. 1996). Members of the SNF2 protein family are generally involved in transcription and cell cycle regulation as well as mitotic chromosome segregation (Carlson and Laurent. 1994). The ATRX helicase domain has been shown to possess ATP-dependant activity that affects the

chromatin state by altering mononucleosome disruption patterns as well as inducing triple-helix DNA displacement (Xue et al. 2003).

A number of additional motifs are present in ATRX, including a nuclear localisation signal (NLS) and a SNF2 conserved motif. Finally, two separate stretches of glutamic acid (E) and glutamine residues (Q) are present in ATRX, these regions are both thought to play a role in protein-protein interactions (Park et al. 2004, Picketts et al. 1996).

ATRX has also been shown to interact with a number of proteins, including: Enhancer of zeste homolog 2 (EZH2) (Cardoso et al. 1998), heterochromatic protein (HP1) (Berube et al. 2002, McDowell et al. 1999), MeCP2 (Nan et al. 2007) and the transcription/apoptosis co-factor, death-domain associated protein (DAXX) (Ishov, Vladimirova and Maul. 2004) (Figure 4.1). In summary, the ATRX structural domains and interacting protein partners suggest a role for this protein as a chromatin remodeler.

In addition to functioning as a chromatin remodeler, ATRX may have a dual role, also participating in cell-cycle regulation. This hypothesis is supported by several lines of empirical evidence, including:

- ATRX is phosphorylated at metaphase and localises to the cytoplasm and decondensed chromatin (Berube, Smeenk and Picketts. 2000)
- ATRX is required for chromosome alignment and meiotic spindle organisation in mouse oocytes (De La Fuente et al. 2004)
- ATRX plays a role in sister chromatid cohesion and chromosome congression during metaphase of mitosis (Ritchie et al. 2008).

ATRX mutations associated with XLMR

To date 113 mutations in *ATRX* have been reported as being causative of XLMR (see supplementary material in Gibbons et al. 2008 for full listing). Generally *ATRX* mutations occur in one of two mutation 'hotspots'. The ADD domain, encoded by exons 7, 8 and the first 300bp of 9, constitutes the first hotspot. Mutations in this region account for 43% of all known mutations to date (Badens et al. 2006). The

second hotspot falls within the helicase domain (encoded by exons 18-29), with mutations in this region contributing to 41% of all reported mutations. However, 33% of *ATRX* mutations occur in exons 18-20 and 26-29, which encode for a portion of the helicase domain. Therefore, collectively mutations in the ADD and part of the helicase domain (exons 18-20 and 26-29) account for 76% of all mutations (Badens et al. 2006). The screening of these regions is therefore a cost effective and logical approach to *ATRX* mutation screening.

The vast majority of *ATRX* mutations are not recurrent and are detected only in single families, with minimal exceptions. These exceptions include the two common *ATRX* mutations, firstly the 536A>G identified in nine families and secondly the 736C>T described in 35 families (Gibbons et al. 2008).

The common 736C>T mutation provides an excellent example of the intra and interfamilial clinical variability between *ATRX* patients. To date, 35 individuals have been identified with this mutation, but the percentage of HbH inclusions (the diagnostic test for alpha thalassaemia) ranges from 0-14% (Gibbons and Higgs 2000). This observation suggests that other factors, which may act either directly or indirectly with ATRX modify the expression of α-globin, thus accounting for this phenotypic variability.

The range of clinical features resulting from *ATRX* mutations are associated with mutations dispersed across the gene, suggesting similar pathophysiological mechanisms for mutations within either of the two domains. This complicates attempts to draw genotype-phenotype correlations with some exceptions. Firstly, it has been observed that mutations which lead to loss of the C-terminal of ATRX are associated with more severe urogenital abnormalities (micropenis and ambiguous genitalia) (Gibbons et al. 2008). Secondly, in a focused study of helicase domain mutations, Baden and colleagues (2006) noted that individuals with mutations affecting this domain had a milder phenotype compared to those patients with ADD domain mutations. This milder phenotype manifests as delayed psychomotor acquisition and absent or mild urogenital abnormalities (Badens et al. 2006).

The pathophysiological basis of *ATRX* mutations

The structure of ATRX (particularly the presence of the ADD and helicase domain) and the protein's interaction with DAXX, EZH2, MeCP2 and HP1, as well as the protein's localisation to heterochromatic regions, strongly supports the hypothesis that ATRX is a chromatin remodeler (as described in Section 4.1.1.1). However, the exact mechanisms by which the deregulation of this physiological process leads to the observed MR phenotype is poorly understood with studies to date giving limited insight.

The hypothesis that the epigenetic state is deregulated in *ATRX* mutation positive patients is supported by changes in the methylation profiles of rDNA genes located within acrocentric chromosomes as well as at Y-specific and subtelomeric repeats (Gibbons et al. 2000). In addition, mutations in the C-terminal of the protein have been shown to attenuate the ability of the protein to localise to promyelocytic leukemia nuclear bodies (PML-NBs). These nuclear bodies include the PML protein in complex with up to 70 other proteins. The PML-NBs are involved in a wide variety of cellular processes one of which includes transcription regulation (Zhong, Salomoni and Pandolfi. 2000). It has been speculated that the reduced localisation of ATRX to the PML-NBs in mutation positive individuals could affect chromatin remodelling and thus transcription repression during neuronal development (Berube et al. 2008)

Full null-*Atrx* mice exhibit embryonic lethality; however, a conditional *Atrx*-knockout in the forebrain of mice led to hypocellularity in the neocortex and hippocampus as well as reduced forebrain size. These anatomical abnormalities are due to increased neuronal apoptosis leading the authors to conclude that *Atrx* is an essential constituent of neuronal development and cell survival (Berube et al. 2005). In addition, the overexpression of *ATRX* in transgenic mice led to a high incidence of embryonic death, neural tube defects and growth retardation. Those mice that did survive to term exhibited seizures, abnormal behaviour, craniofacial abnormalities and prenatal death. These studies showed that ATRX dosage is important for embryonic development, particularly of the brain (Berube et al. 2002).

In support of the dual role for ATRX in transcription and cell-cycle regulation, siRNA *ATRX* knockdown in HeLa cells resulted in aberrant sister chromatid cohesion and chromosome co-segregation at metaphase (Ritchie et al. 2008). This observation was not due to the mislocalisation of HP1, nor were characteristic histone modifications altered, suggesting that ATRX ablation affects mitosis by some other mechanism (Ritchie et al. 2008). In addition, the conditional forebrain of *Atrx*-knockout mice exhibited defects in mitosis in the neuroprogenitor cells of the embryonic forebrain (Berube et al. 2005). These results may explain the abnormal brain development and reduced cortical size associated with ATRX loss.

4.1.2 POSITIONAL CANDIDATE GENE SELECTION: FAMILY FX67

Linkage analysis in family Fx67 led to the identification of an 12Mb disease-associated chromosomal region between markers DXS9896 and DXS993 (Xp21.1-Xp11.4) (Chapter 3). This relatively gene-poor region encompasses 27 known genes (based on human genome assembly 36, March 2006) of which only two have been previously associated with the pathogenesis of XLMR. These XLMR genes, *ATP6AP2* and *TSPAN7* are both associated with mild-moderate non-syndromic MR and are therefore good candidates for mutation detection in NS-XLMR family FX67.

4.1.2.1 Positional candidate gene: ATP6AP2

Phenotypes associated with *ATP6AP2* mutations

To date one XLMR family has been described with a mutation in *ATPase, H+ transporting, lysosomal accessory protein 2* (*ATP6AP2)* (OMIM ***300556**). Hedera et al. (2002) first described a family of seven affected males presenting with mild to moderate MR and epilepsy (designated XMRE). A marked degree of intra-familial variation was also reported, with variable presentation of behavioural problems (aggression, impulsivity, hyperactivity), ataxia and speech delay (Hedera et al. 2002).

ATP6AP2: The gene and its protein

ATP6AP2 was previously mapped to Xp11.4 by genomic sequence analysis (Demirci et al. 2001). The gene consists of nine exons and spans a chromosomal region of approximately 50kb. *ATP6AP2* encodes for a 2.4kb mRNA transcript which is highly expressed in the heart, brain and placenta with intermediate expression in the liver, pancreas and kidney and limited expression in the skeletal muscle and lung (Nguyen et al. 2002).

Upon translation, the *ATP6AP2* mRNA encodes for the renin receptor. The renin-angiotensin system (RAS) is responsible for the maintenance of blood pressure and water-electrolyte balance. Before the identification of the renin receptor it was hypothesised that the primary function of renin was the cleavage of angiotensinogen to angiotensin I (AngI). After a second cleavage stage the active form, angiotensin II (AngII) is produced which exerts its physiological effects on blood-pressure and fluid balance. However, recent studies have shown that the RAS system is involved in additional physiological systems including, and of particular interest here, the CNS where it is responsible for cell proliferation and death, maintenance of neuro-endocrine systems and cognitive properties (McKinley et al. 2003). Some neuronal cell types, particularly astrocytes also express *ATP6AP2* with concomitant production of the renin receptor. Upon activation by renin binding, this membrane-bound receptor was shown to have a fourfold increase in AngII production (Nguyen et al. 2002). In addition this binding led to activation of the mitogen-activated protein (MAP) kinases, ERK1/2, which have been shown to play a role in memory consolidation and long term potentiation (Adams and Sweatt. 2002, Bozon et al. 2003, Nguyen et al. 2002).

ATP6AP2 mutations associated with XLMR

To date, a single family has been reported with an *ATP6AP2* mutation. In this family a c.321C>T, p.D107D was identified by linkage analysis and subsequent positional candidate gene sequencing. This silent mutation disrupts a putative exonic splice enhancer (ESE) site and leads to partial exclusion of exon 4 in 50% of an affected individual's mRNA. This partial exonic exclusion results in a frameshift and ultimately a truncated protein (Ramser et al. 2005).

The pathophysiological basis of *ATP6AP2* mutations

By functional analysis, Ramser et al. (2005) showed that the c.321C>T mutation and the resultant mutant receptor was still able to bind renin and induce catalytic activity to AngI in a way that was comparable to the wild type receptor. However, binding of renin to the mutant receptor resulted in a modest but reproducible reduction in ERK1/2 signalling. It was therefore hypothesised that the XMRE phenotype in this family results from reduced ERK1/2 signalling and concomitant impairment in memory consolidation and long-term potentiation (Ramser et al. 2005).

4.1.2.2 Positional candidate gene: TSPAN7

Phenotypes associated with *TSPAN7* mutations

The second XLMR candidate gene in family Fx67 was *tetraspanin 7* (*TSPAN7*, alias *TM4SF2*) (OMIM ***300096**). Mutations in *TSPAN7* have been identified in four families, including a X:2 female translocation patient who exhibited mild MR and minor autistic features (Zemni et al. 2000). The additional *TSPAN7* mutation positive families all presented with mild to moderate non-syndromic XLMR (Abidi et al. 2002, Zemni et al. 2000).

TSPAN7: The gene and its protein

TSPAN7, located at Xp11.4, spans approximately 20kb of genomic DNA and consists of seven exons. The gene is highly expressed in the hippocampus and cerebral cortex (Zemni et al. 2000).

The *TSPAN7* gene encodes for a member of the transmembrane 4 superfamily, also known as the tetraspanin family of proteins. Members of this protein family are generally cell-surface proteins that control signal transduction of cellular processes such as morphology, invasion, motility, signalling and fusion (Hemler. 2005). In addition, members of this family have been shown to interact with beta-1 integrins, through this interaction it is possible that TSPAN7 controls neurite outgrowth by regulation of the actin cytoskeleton (Zemni et al. 2000).

TSPAN7 mutations associated with XLMR

The first *TSPAN7* mutation was identified in an isolated female MR patient, this individual exhibited loss of gene expression due to an X:2 translocation (Zemni et al. 2000). Subsequent to the identification of this XLMR gene, analysis of probands from 33 XLMR families led to the identification of two additional disease-causing mutations (Zemni et al. 2000). Finally, Abidi et al. (2002) reported a 2bp deletion (564delGT) which led to a frameshift and premature truncation of the protein at amino acid 192 (Abidi et al. 2002).

The pathophysiological basis of *TSPAN7* mutations

Little is known about the pathogenic mechanisms that are induced by *TSPAN7* mutations. All mutations lie in the larger of two extracellular loops responsible for protein-protein interactions, but little is known of the protein substrates or disease mechanism (Abidi et al. 2002, Hemler. 2005, Zemni et al. 2000). Given the neuronal localisation of TSPAN7, as well as its putative interaction with beta-1 integrins, it has been hypothesised that TSPAN7 mutations lead to reduced regulation of the actin cytoskeleton with concomitant effects on neurite development and outgrowth (Zemni et al. 2000).

4.2 METHODS

4.2.1 MUTATION DETECTION IN *ATRX*: FAMILY XMR2

4.2.1.1 Clinical presentation of patients in family XMR2

The seven affected males in family XMR2 presented with a syndromic form of XLMR with a broad spectrum of clinical features (Table 4.1). Three of the seven affected individuals in family XMR2 are deceased; two succumbed to accidental deaths and a third died at the age of 38 years from pneumonia following recurrent chest infections and unexplained weight loss. The surviving men have an age range of 31 to 40 years.

Table 4.1 The clinical presentation of the seven affected males in family XMR2

Clinical presentation	**# of affected individuals with feature (%)**
Level of MR	
Severe	4 (57)
Moderate	3 (43)
Craniofacial features	
Microcephaly	6 (86)
Narrow bi-frontal diameter	3 (43)
Hypertelorism	7 (100)
Prognathism (more pronounced in puberty)	5 (71)
Carp-shaped mouth	2 (29)
Small, posteriorly-rotated ears	5 (71)
Widows peak	7 (100)
Premature greying of forelock	4 (57)
Speech	
Absent	3 (43)
Single words	3 (43)
Short sentences	1 (14)
Other	
Aggressive and disruptive behaviour	3 (43)
Epilepsy (onset after puberty)	3 (43)
Hypotonia (in infancy)	7 (100)
Short stature (below 3rd centile)	7 (100)
Gastro-oesophageal reflux	2 (29)
Recurrent chest infections	3 (43)

The clinical presentation of affected individuals was characterised by intrafamilial variation, particularly with respect to patients' neurocognitive abilities (Table 4.1). For example, only one affected male was able to speak in short sentences as compared to the remainder of the patients who could only speak short words or had no speech at all. Interestingly, this individual was also the only affected member of this family to have a normal head circumference. Similarly, there was a large age range (3.5–7.5 years) within which affected individuals learnt to walk.

Detailed neuroimaging investigations have not been performed in the affected individuals of this family. Only one patient underwent an MRI scan which showed features of cerebral atrophy and periventricular leucomalacia but his premature birth was likely to have contributed to the findings. It should also be noted that no genital abnormalities were detected in this family. Finally, only a single affected individual in this family (XMR2.5) underwent biochemical evaluation for HbH inclusions, this test was shown to be negative, a diagnosis of α-thalassaemia was therefore excluded.

4.2.1.2 mRNA isolation and cDNA synthesis

Mutation detection analyses were performed in the *ATRX* mutation 'hotspots' i.e. exons 7-9 (hotspot 1 in the ADD domain), exons 18-20 (hotspot 2 in the helicase domain) and exons 26-29 (hotspot3 in the helicase domain). In order to exclude investigation of intronic DNA regions, mRNA and subsequently synthesised cDNA, was used for this mutation screening. MRNA was isolated from an affected individual (XMR2.5) using the PAXgene™ Blood RNA System [PreAnalytix] according to manufacturer's instructions. Subsequently, complete cDNA synthesis was performed using the iScript™ cDNA Synthesis Kit [Bio-Rad] as per manufacturer's instructions.

4.2.1.3 cDNA PCR amplification of ATRX hotspots

PCR primers were designed in order to amplify the three mutation 'hotspots' of *ATRX* (GenBank accession no. **NM_000489.2**) from the synthesised cDNA (Appendix 4A). PCR amplifications were performed in a 25µl reaction containing 400ng of patient's cDNA, 10µM of each forward and reverse primer, 2.5µM dNTP's [Bioline], 1× Go Taq buffer [Promega] and 1U Go Taq polymerase [Promega]. The PCR profile consisted of an initial denaturation step at 95ºC for 5 min, followed by 30 cycles at 95ºC for 1min, 57 ºC for 1min, and 72ºC for 1min. A final elongation at 72ºC for 5 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation]. PCR products were visualised by EtBr staining subsequent to electrophoresis on a 3% agarose gel alongside a 100bp molecular weight marker [Fermentas].

4.2.1.4 ATRX genomic DNA PCR amplification

In order to elucidate the genomic DNA context of a cDNA mutation corresponding to exon 26, PCR primers were designed to flank the 5' intronic region of exon 26 (intron 25). This primer set consisted of the forward primer: 5' ttctgggatagtctctgtcc 3' and reverse primer: 5' ccaatttcctctgccattcg 3'. PCR was performed as above (Section 4.2.1.3) with the exception of the use of 100ng of patients' genomic DNA as opposed to cDNA. The PCR profile consisted of an initial denaturation step at 95ºC for 5 min, followed by 30 cycles at 95ºC for 30s, 55ºC for 30s, and 72ºC for 30s. A

final elongation at 72ºC for 5 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation]. PCR products were subject to the same electrophoresis protocol as above (Section 4.2.1.3).

4.2.1.5 DNA sequencing analysis of DNA amplicons

Sanger DNA sequencing was performed in order to identify the putative underlying *DNA* mRNA/DNA sequence variations. The total volume of DNA products generated from amplification of the cDNA mutation 'hotspots' and the exon 26 genomic DNA region were electrophoresed on a 1% agarose gel and the bands excised. Purification of the fragments was achieved using the Qiagen™ Extraction Kit according to the standard manufacturer's protocol.

The direct DNA sequencing reaction was performed in a 10µl final reaction-volume containing 500ng of purified DNA template, 10µM of either forward/ reverse primer, 2µl Big Dye terminator mix [Applied Biosystems] and 1× Big Dye terminator sequencing buffer [Applied Biosystems]. The sequencing profile entailed an initial denaturation at 94ºC for 5 minutes followed by 25 cycles of 94ºC for 20s, 50ºC for 10s and 60ºC for 4min. The Applied Biosystems GeneAmp® DNA system 9700 thermal cycler was used for this amplification.

Purification of DNA sequencing products was performed using ethanol precipitation. All DNA sequencing products were electrophoresed on the ABI Prism® 3100 Genetic Analyser and analysed using the Sequence Analysis version 3.7 Software [Applied Biosystems]. BioEdit Sequence Alignment Editor (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) was used to align the resultant DNA sequence to the reference sequence (Homo Sapiens, March 2006 version).

4.2.2 MUTATION DETECTION IN *TSPAN7* AND *ATP6AP2*: FAMILY FX67

4.2.2.1 Clinical information of family FX67

Family FX67 consisted of three affected brothers who presented with NX-XLMR. The only additional clinical feature was epilepsy, present in all three affected males.

4.2.2.2 PCR amplification of ATP6AP2 and TSPAN7 exons

PCR primers were designed for the nine *ATP6AP2* exons (GenBank **NM_005765.2**) and the seven protein-coding *TSPAN7* exons (GenBank **NM_004615.2**) (Appendix 4A). PCR amplification for all exons of *ATP6AP2* and *TSPAN7* were performed as described in Section 4.2.1.4 with the following exceptions:

- The primer annealing temperature for *ATP6AP2* exons 1, 2, 3, 5-9 was 56ºC.
- The primer annealing temperature for exons 3 and 8 of *ATP6AP2* and exons 2-7 of *TSPAN7* was 60ºC
- The primer annealing temperature of *TSPAN7* exon 1 was 64ºC and Failsafe buffer J [Epicentre Technologies] replaced the 2.5µM dNTP's [Bioline] and 1× Go Taq buffer [Promega].

4.2.2.3 DNA sequencing of ATP6AP2 and TSPAN7 exons

DNA sequencing was performed as before using *ATP6AP2* and *TSPAN7* specific primers (Section 4.2.1.5).

4.2.2.4 X-inactivation analysis of unaffected mother Fx67.1

Up to 50% of XLMR carrier females exhibit skewed X-inactivation (i.e. >>80:20) (Plenge et al. 2002). Therefore, the presence of skewed X-inactivation in the unaffected mother (Fx67.1) of family Fx67 would be highly indicative of the disorder in this family indeed being attributed to an X-linked mutation. Consequently, in order to better estimate the possibility of the condition being X-linked in this trio of brothers, an X-inactivation assay was performed before screening novel XLMR candidate genes. To assess the X-inactivation pattern of individual Fx67.1, the methylation status of a CpG island located within the variable number of tandem repeats (VNTR) region of the *MAOA* gene was analysed as previously described (Section 3.2.4.2)

4.3 RESULTS

4.3.1 MUTATION DETECTION IN *ATRX*: FAMILY XMR2

4.3.1.1 cDNA sequencing analysis of ATRX mutation hotspots

DNA sequencing analysis of the first mutation hotspot (exon7-first 300bp of exon 9) as well as the second (exons 18-20) showed no alteration in the cDNA sequence of the affected individual (XMR2.5). However, the PCR product corresponding to the third mutation hot spot (exons 26-29) showed a reduction in size as compared to the control sample when analysed by agarose gel electrophoresis (Figure 4.2). Sequencing analysis confirmed a 66bp deletion in the cDNA of the affected individual (Figure 4.3). This cDNA deletion correlates to the first 66bp of exon 26 and was designated r.5957_6022del. The deletion of these 66 nucleotides results in a loss of 22 amino acids, p.1987_2008del as well as an alteration of the preceding amino acid, p.1986S>N.

4.3.1.2 Genomic DNA sequencing analysis of r.5957_6022del.

In order to characterise the genomic DNA context of the r.5957_6022del, DNA sequencing of the intron/exon boundaries of exon 26 was performed. This analysis revealed a 24bp deletion (as compared to the 66bp mRNA deletion) within exon 26 and was designated c.5987_6011del (Figure 4.3). These findings show that the genomic deletion causes aberrant splicing of exon 26, leading to the exclusion of the 5' end of the exon (Figure 4.3). The c.5987_6011del segregated with MR as it was identified in all affected individuals of family XMR2 as well as the carrier mothers (Figure 4.4).

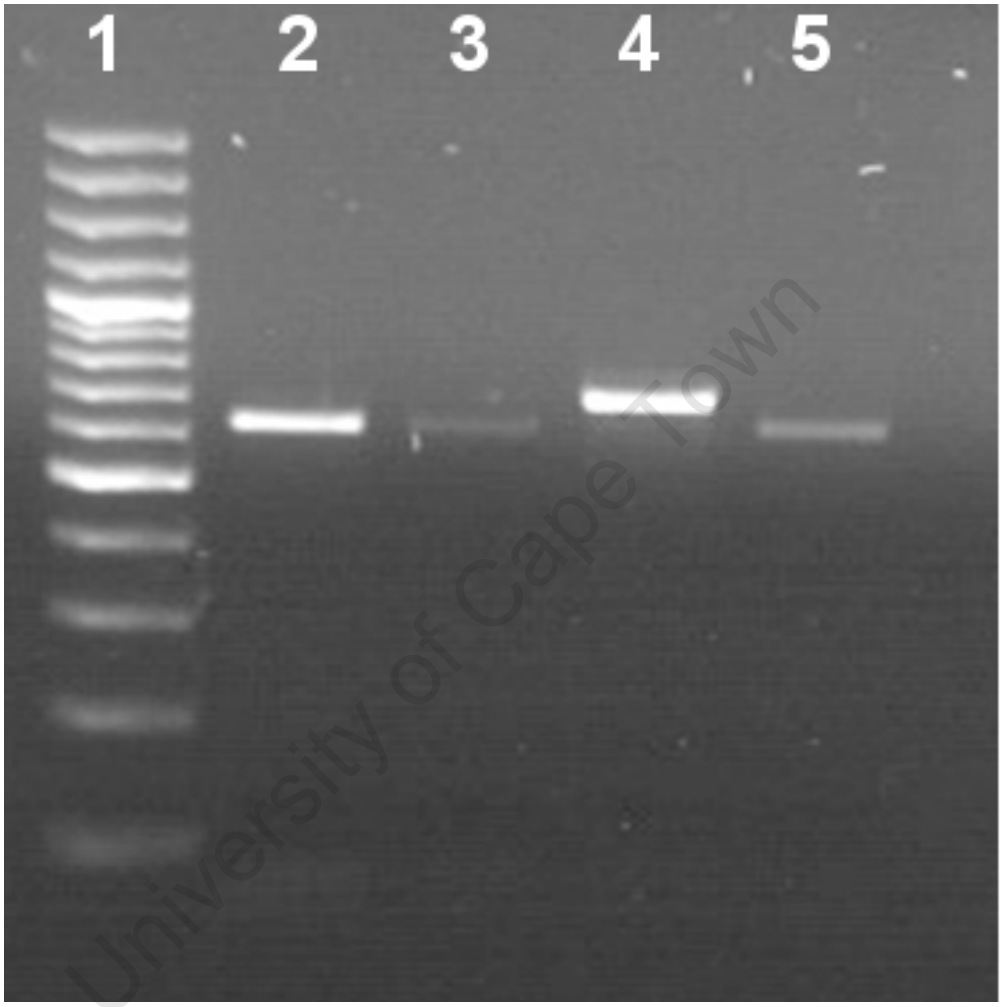


Figure 4.2 Results of cDNA PCR amplification for the two *ATRX* helicase domain 'hotspots'. PCR products were separated by gel electrophoresis (2% agarose gel) and visualised with EtBr. Lane 1 shows the 100bp molecular weight marker (GeneRuler™ 100bp DNA Ladder Plus [Fermentas]**)**. The second and third lanes show the results of PCR amplification for the first helicase domain hotspot. A band of the correct size (626bp) was obtained for both the control (Lane 2) and the patient (XMR2.5) samples (Lane 3). The intensity of the band is lower in the patient sample (XMR2.5), most likely indicative of reduced levels of *ATRX* mRNA transcript. The final two lanes represent the second helicase domain mutation hotspot demonstrating the control PCR product of 687bp (Lane 4) and the 621bp PCR product in XMR2.5 (Lane 5) demonstrating the 66bp deletion (r.5957_6022del) in this XLMR patient. Once again, the intensity of the band (XMR2.5 - Lane 5) illustrates a reduced transcript level as compared to the control (Lane 4), suggesting reduced mRNA transcript levels.

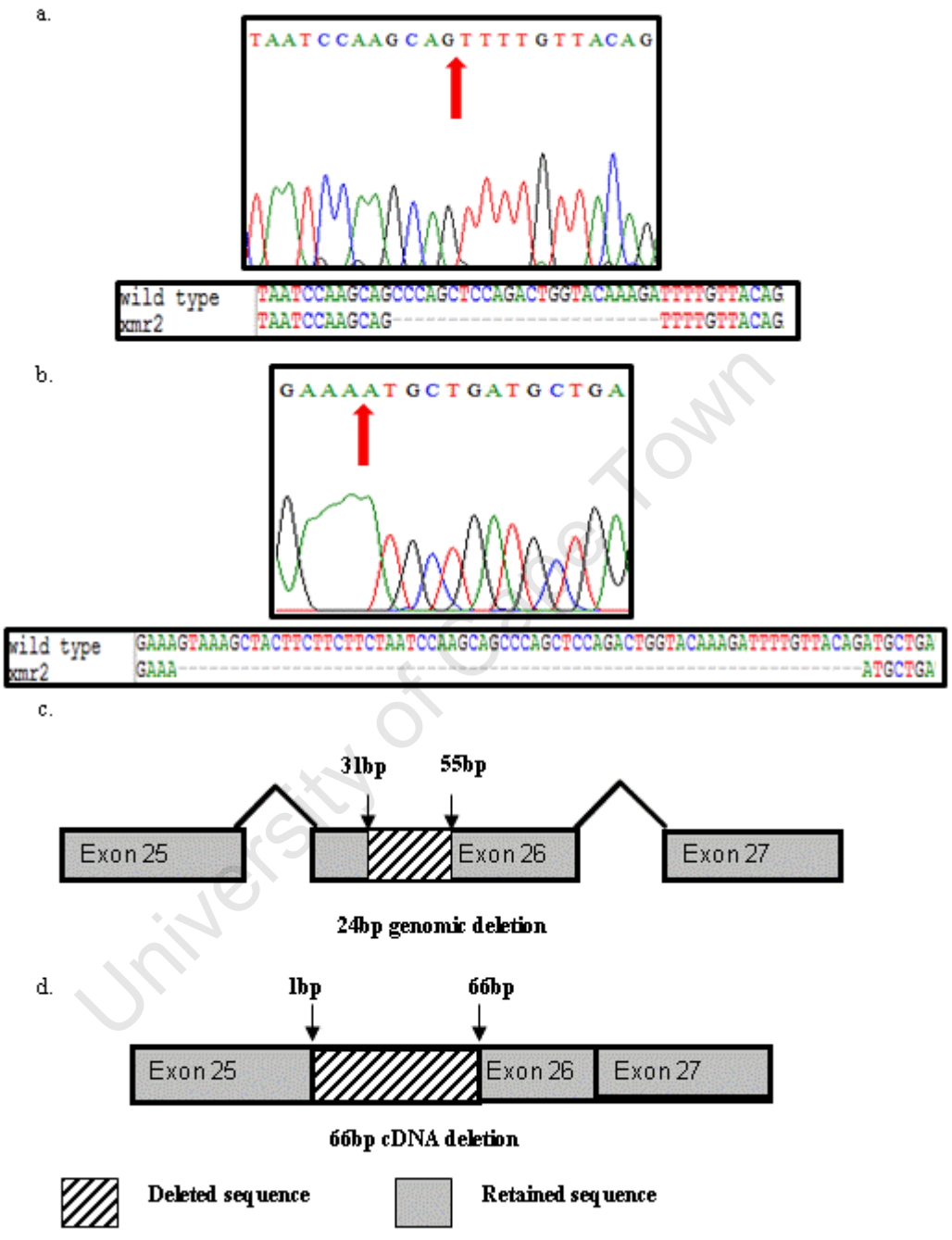


Figure 4.3 The mRNA mutation (r.5956_6022del) and corresponding genomic DNA deletion (c.5987_6011del) in patient XMR2.5. (a) Sequencing of genomic DNA of *ATRX* exon 26 showing the sequence flanking the 24bp deletion (c.5987_6011del), red arrow indicates nucleotides flanking the deletion site. (b) Sequencing of cDNA of the DNA encoding for the second chromatin remodelling domain showing the 66bp deletion (r.5956_6022del), red arrow indicates nucleotides flanking the deletion site. (c) location of the 24bp genomic deletion (c.5987_6011del) in exon 26 spanning base pairs 31-55, also illustrating the retention of the donor/acceptor splice sites. (d) The resultant 66bp cDNA deletion (r.5956_6022del), spanning the first 66bp of exon 26 due to an aberrant splicing mechanism caused by the 24bp genomic deletion.

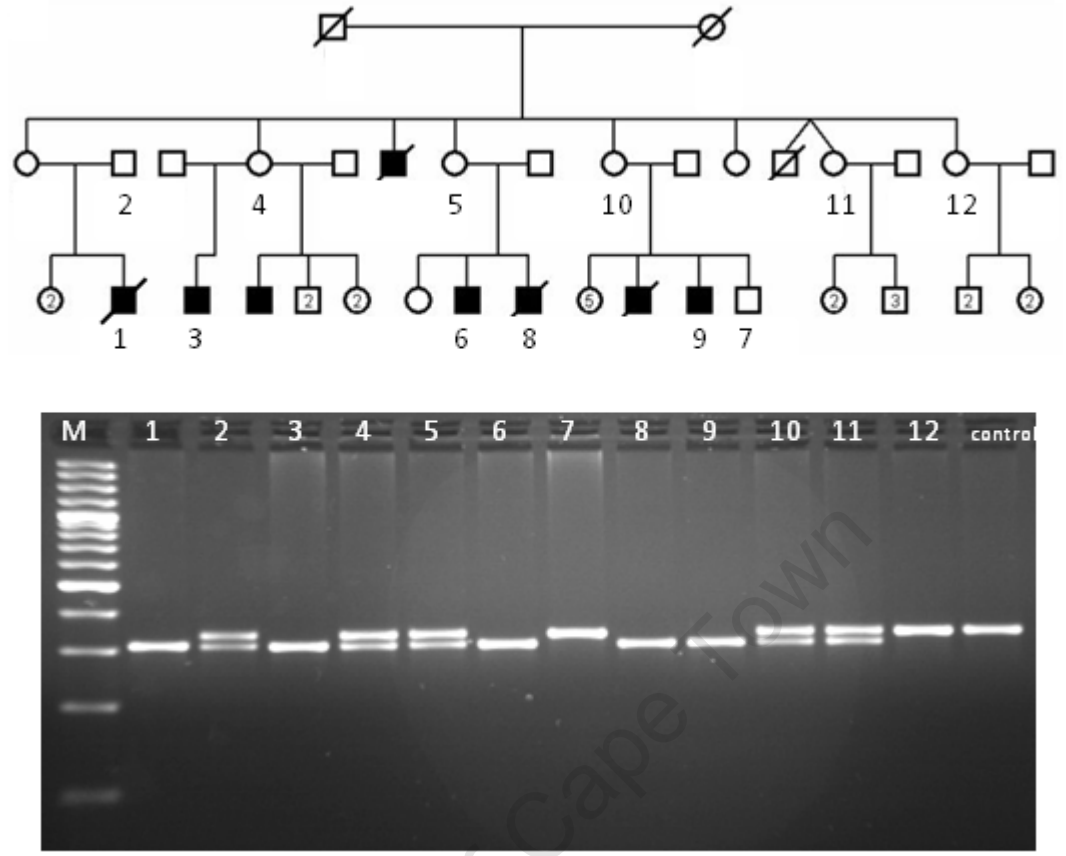


Figure 4.4 Co-segregation of the 24bp genomic DNA deletion (c.5987_6011del) with MR in family XMR2 (a) Pedigree of *ATRX* mutation positive family. (b) Agarose gel (3%) electrophoresis (visualised with EtBr) subsequent to PCR amplification of the intron/exon boundary of exon 26. The first lane shows the 100bp molecular weight marker (GeneRuler™ 100bp DNA Ladder Plus [Fermentas]**)**. Subsequent lanes show the deletion segregating in the family with amplicon sizes of 335bp (normal allele) and 311bp (mutant allele - c.5987_6011del), all lane numbers correspond to numbering system of the Fx444 family pedigree in (a).

4.3.1.3 Analysis of ESE consensus sites within the c.5987_6011del

The c.5987_6011del identified in family XMR2, resulted in aberrant exon 26 splicing (exclusion of the first 66 bp of exon 26). Given that the c.5987_6011del did not disrupt the acceptor/donor sites of exon 26, this alternate splicing was due to a disruption of an alternative consensus splicing motif. Therefore, the integrity of exonic splice enhancer (ESE) sites was assessed using the programme ESE finder (http://rulai.cshl.edu/cgi-bin/tools/ESE/). This in silico prediction tool demonstrated that the c.5987_6011del led to the disruption of a Srp40 and SC35 consensus sequence (Figure 4.5). It was probable that the disruption of one or both of these sites leads to the partial exonic exclusion in the *ATRX* transcript in XMR2 affected individuals.

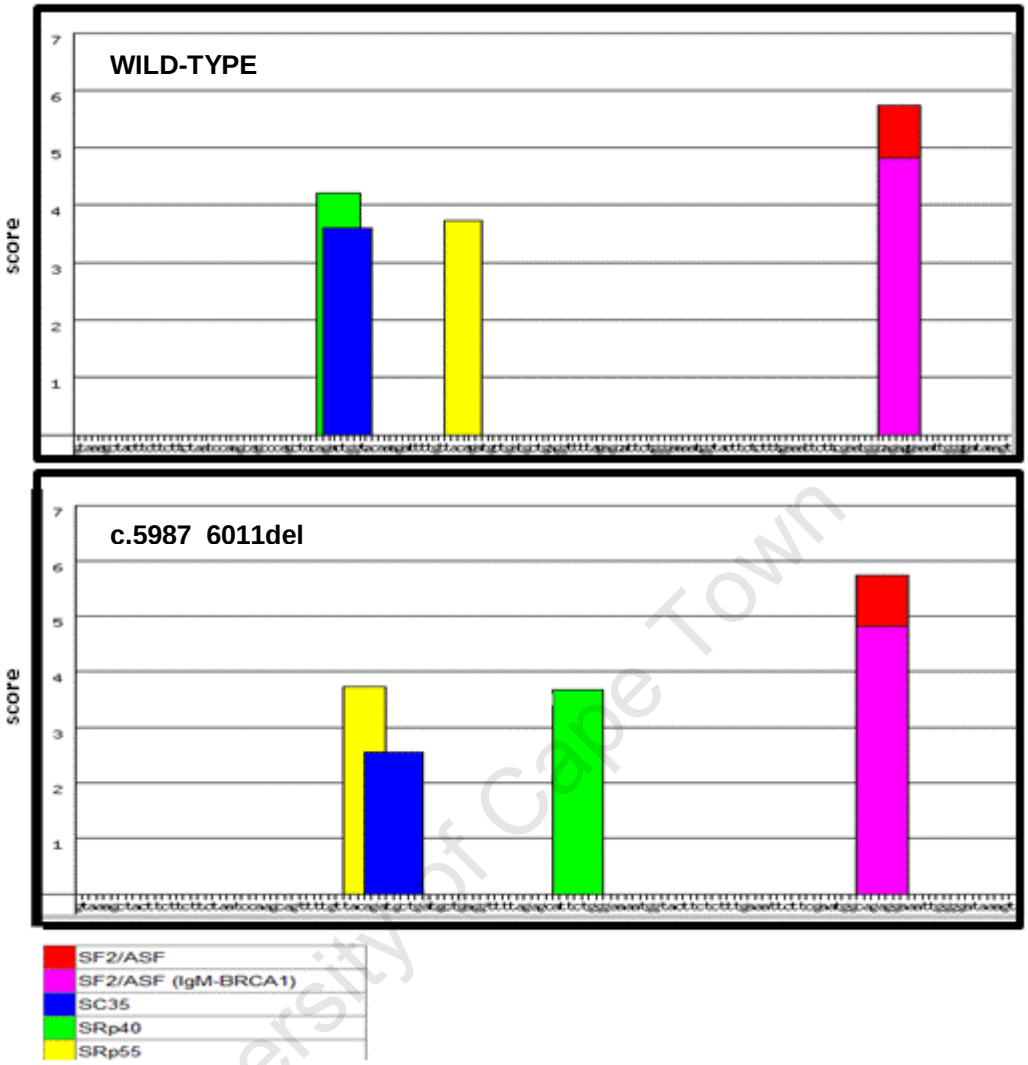


Figure 4.5 The distribution of the highest-scoring consensus sequence locations for each of the five ESE sites in the wild type as well as the *ATRX* c.5987_6011del in family XMR2, as predicted by the ESEFinder tool. The illustrations above show the distribution of the predicted ESE binding sites, according to exon 26 DNA sequence (y-axis), as well as the probability scores of each of these sites (x-axis). The results of this in silico predictive tool, suggested that the c.5987_6011del disrupts a Srp40 (green) and SC35 (blue) consensus sequence.

4.3.2 MUTATION DETECTION IN *ATP6AP2* AND *TSPAN7*: FAMILY FX67 POSITIONAL CANDIDATE GENES ANALYSIS.

4.3.2.1 Genomic DNA sequencing analysis of ATP6AP2

No DNA sequence variations were detected in any of the nine *ATP6AP2* exons evaluated by genomic DNA sequencing.

4.3.2.2 Genomic DNA sequencing analysis of TSPAN7

No DNA sequence variations were detected in exons 2-7 of the *TSPAN7* gene by genomic DNA sequencing. However, two variants were detected in the 5'UTR in all affected individuals (Fx67.2/3/4). The first was a transition, designated c.-125G>A and the second a 3bp deletion designated c.-28_30delCCG.

The c.-125G>A transition was described in the Ensembl database (Ensembl release 52, December 2008) as a SNP, but there was no report in dbSNP (NCBI). The Ensembl record (**ENSSNP3520788**) shows no population or linkage disequilibrium data. Individual genotypes are available from Craig Venter's sequenced genome, this individual has the A variant. Given the normal cognitive function in this individual it was concluded that the c.-125G>A variant does not contribute to the disease aetiology in this family.

The 3bp deletion (c.-28_30delCCG) was present in dbSNP (NCBI), accession number: **rs3833408**. There was no population, linkage disequilibrium or individual data available for the variant and it has not been validated by independent methods or multiple submissions. At date of submission of this dissertation this variant has no known clinical associations and its role in the pathogenesis of XLMR in family FX67 remains unknown.

4.3.2.3 X-inactivation analysis in unaffected female Fx67.1

In order to ascertain the X-inactivation pattern in the affected boys' (Fx67.2/3/4) mother (Fx67.1), X-inactivation analysis was performed at the *MAOA* VNTR locus. From these analyses it was established that the X-inactivation ratio was 72:28 (Figure 4.6). The allele (and therefore X-chromosome) which was preferentially inactivated in Fx67.1 (318bp *MAOA* VNTR allele) was the same allele which was transmitted to all three affected boys. These results indicate that there was mild preferential inactivation of the 'affected' X chromosome (assuming X-linked inheritance of MR).

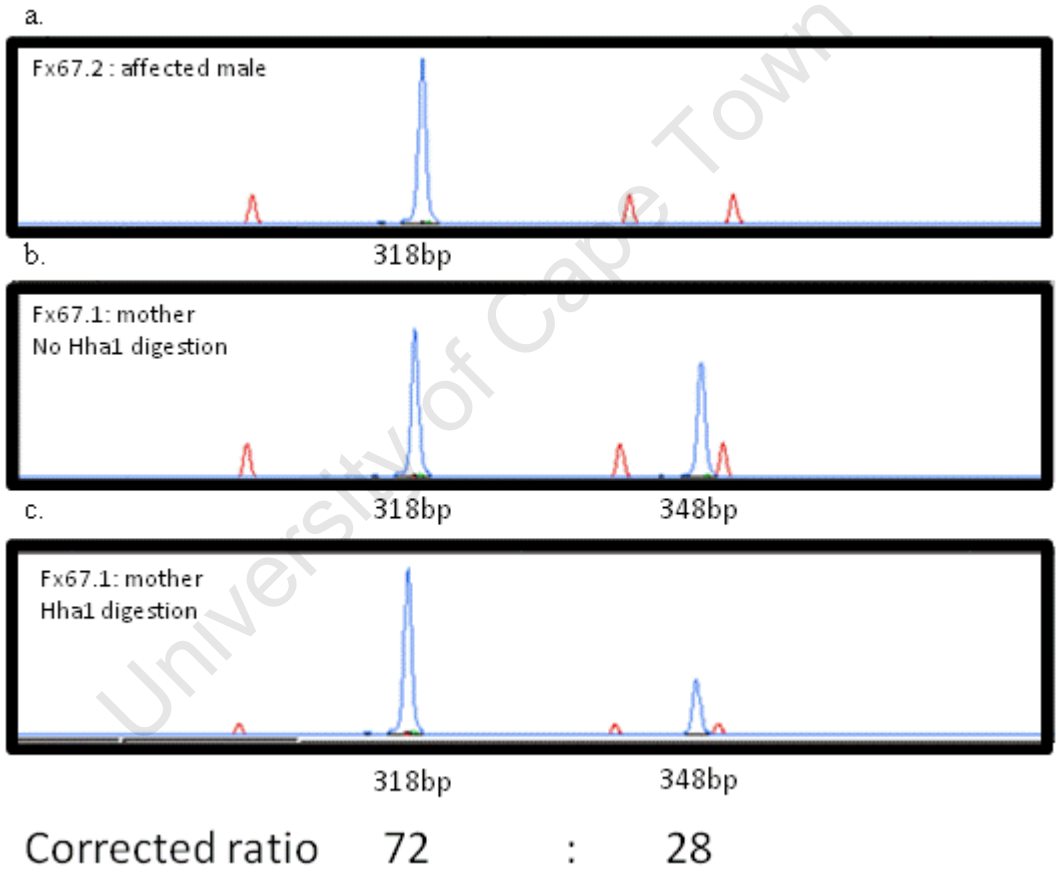


Figure 4.6 The X-inactivation pattern in the unaffected mother (Fx67.1) of the three male-sibs presenting with NS-XLMR in family Fx67 (i.e. Fx67.2/3/4). (a) The electropherogram generated by size-based capillary electrophoresis illustrating the *MAOA* VNTR allele in affected individual Fx67.2. All affected brothers carry the 318bp allele (blue) at this locus. The 500bp Rox size standard [Applied Biosytems] is illustrated in red (b) The electropherogram illustrating the intensity of the two *MAOA* VNTR alleles (318bp and 348bp) (blue) prior to digestion of the unaffected mother's DNA sample with the methylation-specific enzyme HhaI. There waa evidence in this electropherogram of preferential amplification of the smaller (318bp) allele (blue) as compared to the larger (348bp) allele (blue). (c) The electropherogram illustrating the intensity of the two *MAOA* VNTR alleles (318bp and 348bp) (blue) after digestion of the unaffected mother's DNA sample with the methylation-specific enzyme HhaI. Once the allele-specific PCR amplification bias was accounted for it was concluded that the X-inactivation ratio was 72:28.

4.4 DISCUSSION

Mutation detection in positional candidate genes subsequent to linkage analysis is an important step in identifying the underlying molecular defect in a family with a genetic disorder. In this study, linkage analysis identified a 10Mb and 11Mb minimal critical interval in XLMR families XMR2 and FX67 respectively (Chapter 3). Positional candidate genes were selected from these critical intervals and mutation detection was performed, in order to ultimately identify mutation positive patients applicable for epigenetic profile analyses.

The critical chromosomal interval in family XMR2 idenitied in this study encompassed approximately 70 genes, including nine known XLMR genes (*DLG3, MED12, NLNGN3, SLC16A2, KIAA2022, ATRX, ATP7A, IAP* and *PGK1*). Subsequent to a thorough clinical review of affected individuals in family XMR2, the overall familial clinical picture was more suggestive of *ATRX* being associated with XLMR in this family. Therefore, *ATRX* was selected as our primary candidate gene and mutation detection via direct sequencing of the *ATRX* mutation hotspots was performed. These investigations identified a novel splicing mutation, c.5987_6011del in the ATRX helicase domain.

The 24bp genomic deletion, c.5987_6011del results in a larger, 66bp deletion in the mRNA transcript (r.5957_6022del). Interestingly, this mRNA mutation has been described previously in two independent reports (Gibbons and Higgs. 2000, Gibbons et al. 2008). However, the underlying genomic mutation in these reports was different to the deletion described here (c.5987_6011del). The two previously reported mutations include an intronic c.5957-2A>G splicing variant and a nonsense mutation, c.6003G>A, which was predicted to give rise to premature truncation (p.W2001X) (Gibbons and Higgs. 2000, Gibbons et al. 2008). Given that these three different genomic DNA mutations (c.5987_6011del, c.5957-2A>G, c.6003G>A) result in the same larger mRNA deletion, it was likely that these genomic mutations induce the same aberrant splicing event, ultimately resulting in partial exonic exclusion. The ensuing aberrantly-spliced transcript encodes a protein with 22 amino acids absent from the helicase domain. Despite the predicted deleterious effect on ATRX function, this splicing variation may in fact be acting as a form of

phenotypic rescue as an ATRX null allele is expected to be lethal (Garrick et al. 2006). These three genomic mutations could all encode for a truncated ATRX protein. Firstly, the c.6003G>A mutation encode for a p.W2001X truncating mutation. Secondly, the c.5957-2A>G, was predicted to lead to non-recognition of the AG acceptor site, thus putatively leading to whole exon 26 skipping (or cryptic activation of a downstream acceptor site). This reading frame alteration could result in a frameshift, ultimately leading to premature ATRX truncation. Lastly, the c.5987_6011del identified in this study, was predicted to give rise to a frameshift mutation which would also lead to premature truncation 56 amino acids downstream (p.1987fsX56). Therefore, we hypothesise that the observed aberrant splicing was due to interruption of a common exonic ESE site (in the cases of c.5987_6011del and c.6003G>A) and alternative recognition of a donor site (c.5957-2A>G) resulting in the activation of a downstream cryptic ESE site.

ESE sites are regions of conserved exonic sequence which are believed to act as binding sites for Ser/Arg (SR) rich proteins. These SR proteins are responsible for exon definition through protein-protein interaction with other components of the splicesome, essential for correct pre-mRNA splicing (Figure 4.7) (Cartegni, Chew and Krainer. 2002).

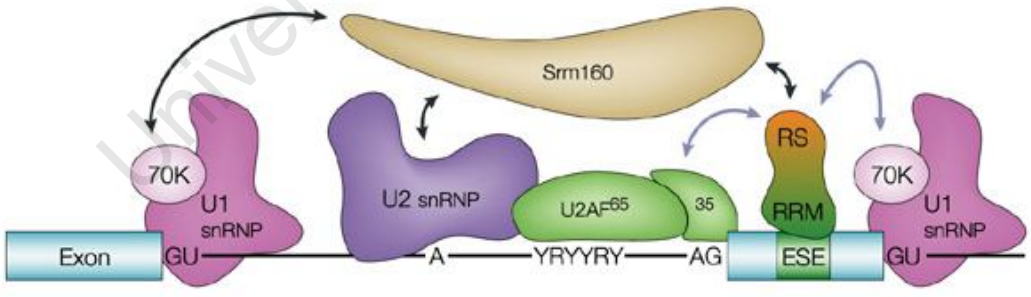


Figure 4.7 A theoretical model for exonic splicing through the formation of the splicesome. SR proteins bind to the ESE consensus sequence through its RNA-recognition motifs (RRM). Through its Arg/Ser rich (RS) domain the SR protein also interacts with U2AF (35 subunit) and U1 snRNP (70K subunit) which bind to the acceptor AG and donor GU sites respectively. These components then either directly or indirectly recruit the remainder of the splicesome complex (from Cartegni, Chew and Krainer. 2002).

The programme ESE finder (http://rulai.cshl.edu/cgi-bin/tools/ESE/) uses weighted matrix values and consensus sequences for each of the four SR proteins (SF2/ASF, SC35, SRp40, SRp55) to computationally predict functional ESE sites (Cartegni et al. 2003). The use of this prediction tool showed that the c.5987_6011del and c.6003G>A mutations lead to the interruption of a common predicted SRp40 consensus sequence located at nucleotides 42-48 of *ATRX* exon 26 (Figure 4.8). This interruption could lead to the activation of a downstream cryptic SRp40 site located at exon 26 nucleotides 91-98 (numbered according to the wildtype transcript). To further support this hypothesis there is evidence suggesting that the SR rich proteins which bind at ESE sites interact with the U2 auxiliary factor 35 ($U2AF^{35}$) which binds at the AG acceptor site (Figure 4.8) (Cartegni, Chew and Krainer. 2002). The mutation at c.5957-2A>G could lead to the recognition of a cryptic AG acceptor site (located 38bp upstream of the cryptic SRp40 site) (Figure 4.8). Assumption of this model would provide a plausible explanation of the common mRNA mutation despite independent genomic DNA mutations.

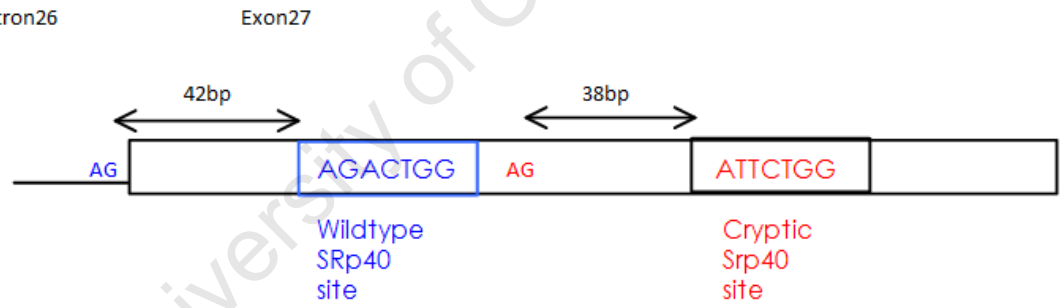


Figure 4.8 Illustration of the putative alternative splicing mechanism by which the three genomic mutations give rise to the same 66bp mRNA deletion (r.5957_6022del). First, illustrated in blue is the wildtype SRp40 consensus ESE binding site within *ATRX* exon 26. The consensus sequence at this site is disrupted by the c.6003G>A mutation, becoming AGACTGA (hence not recognised by SRp40), and the site is deleted completely by the c.5987_6011del mutation. These mutations could lead to the activation of the downstream cryptic SRp40 site (illustrated in red). Under normal conditions the AG acceptor site is recognised by $U2AF^{35}$ which in turn interacts with SRp40 to form part of the splicesome. However, the c.5957-2A>G abolishes this site leading to the cryptic recognition of the AG site and subsequent interaction with the downstream SRp40 ESE. In this way all three mutations are hypothesised to make use of the downstream SRp40 site to induce mutation skipping and subsequent phenotypic rescue.

It should, however, be noted that this proposed model presents merely a theoretical explanation, in order to test this hypothesis an in vitro splicing assay could be performed. This would entail the cloning of cDNA PCR products from all three patients (or site-directed mutagenesis) into an exon trapping plasmid with subsequent

assessment of the resultant mRNA transcripts. Double knock-out mutants in both the wildtype and cryptic SRp40 sites would further elucidate the validity of this theoretical hypothesis.

Previously Wada and colleagues (2006) demonstrated that the disruption of an ESE site by a single base pair change (c.370G>T) results in the exclusion of several exons in *ATRX* (Wada et al. 2006). While this missense mutation (p.G124C) was located outside the ADD domains, it lies in a highly conserved region of 36 amino acids (Park et al. 2004). Therefore, it was likely that the region encompassed by this mutation undergoes the same splicing-induced phenotypic rescue as described above. Finally, it has been previously suggested that phenotypic rescue mechanisms such as those described here are present in all truncating mutations upstream of the helicase domain (Gibbons et al. 2008).

The clinical presentation of affected individuals in family XMR2 was characterised by intrafamilial variability especially with respect to the patients' neurocognitive abilities. It would be interesting to conduct quantitative mRNA investigations in the surviving affected males in order to investigate the degree of phenotypic rescue as compared to the severity of MR and additional clinical features. Unfortunately, at this stage, collection of fresh blood samples for RNA analyses in all surviving patients has not been feasible; therefore, it has not been possible to perform these investigations.

The mutant mRNA transcript (r.5956_6022del), when translated, results in an ATRX protein lacking 22 amino acids (p.1987_2008del) and an altered amino acid distal to the deletion (p.1986S>N). This mutation resides in the functionally significant C-terminal helicase domain of the ATRX protein and deletes a number of highly conserved residues (Park et al. 2004). It has been shown that mutations in the helicase domain affect the ATPase activity of the ATRX protein. A study by Tang and colleagues (2004) demonstrated reduced activity in recombinant ATRX carrying missense mutations involving non-conserved amino acids in the helicase domain. This led the authors to hypothesis that the deletion of highly conserved domains in this region would be lethal as the ATPase 'dead' mutant completely abolished ATPase activity (Tang et al. 2004). The non-lethal c.5987_6011del mutation

described here involves the deletion of a highly conserved region and would seem to contradict this statement. However, this mutation was not located in the catalytic sites and could therefore only impair ATPase activity rather than completely abolish function. Therefore, the findings of this study suggest that the pathogenic effect of amino acid changes in the helicase domain is dependent on the location as well as the conservation of the amino acid in question.

This study has identified a novel c.5987_6011del as the mutation underlying a syndromic form of XLMR in a South African family. The clinical presentation of affected individuals includes severe MR, microcephaly and hypotonic facial features, but excluded α-thalassaemia. While 90% of individuals with *ATRX* mutations present with α-thalassaemia, studies suggest that the absence of this haematological feature should not be an exclusion factor in the decision to conduct *ATRX* mutation screening (Villard et al. 1999). Villard and colleagues (1999) propose that the most valuable diagnostic criteria for screening of the *ATRX* hotspots are MR, microcephaly and hypotonic facial features (Villard et al. 1999). In addition, the large percentage of α-thalassaemia in *ATRX* mutation positive patients may be a reflection of selection bias towards patients exhibiting α-thalassaemia. Therefore, unbiased mutation screening in affected individuals presenting with only clinical features outlined by Villard and colleagues, may result in a reduction in the percentage of *ATRX* mutation positive patients presenting with alpha thalassaemia. Certainly, the results of this study would support this screening strategy. Screening of larger clinically stratified cohorts will lend an indication to the feasibility of including *ATRX* mutation 'hotspot' testing in the diagnostic protocols for MR patients, particularly in a developing country such as South Africa.

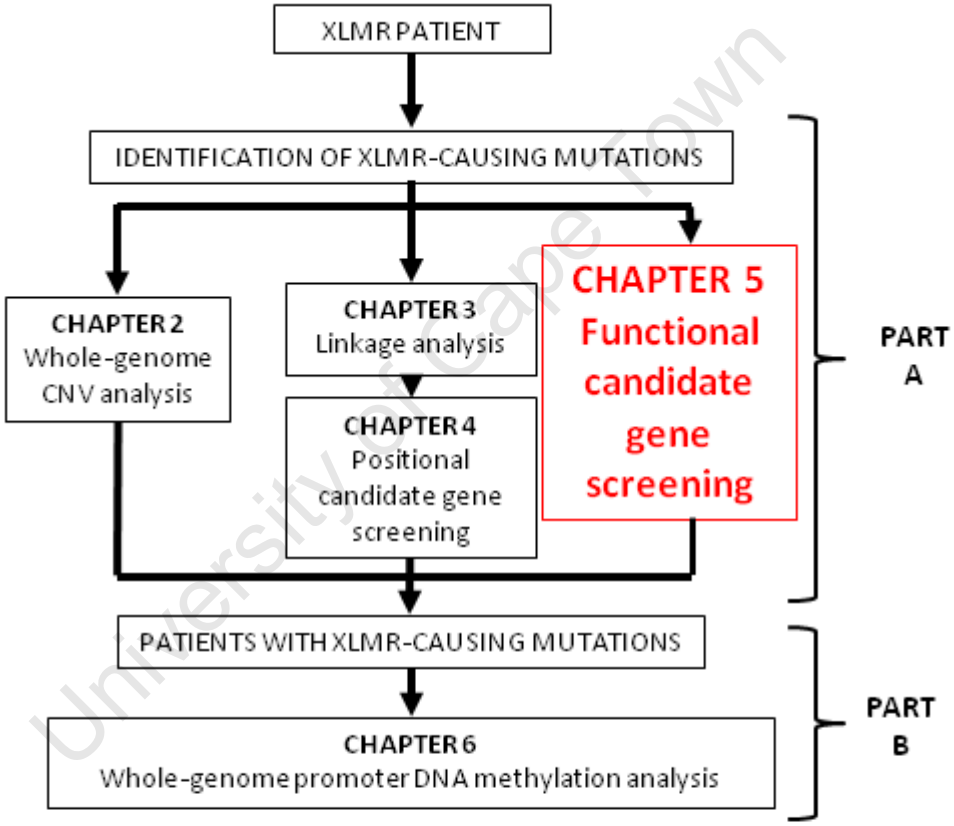
The critical chromosomal interval identified in family FX67 spanned an 11Mb chromosomal region (Xp21.1-Xp11.4), which encompassed 27 genes, two of which are known XLMR genes, *ATP6AP2* and *TSPAN7*. Mutation detection in both of these positional candidate genes revealed no disease-causing mutations. Two SNPs were identified in exon 1 of the *TSPAN7* gene; these alterations are most likely not associated with the disease-phenotype in this XLMR family owing to their occurrence in the general population.

In addition to the two known XLMR genes, the XLMR-associated chromosomal interval in family Fx67 encompasses an additional 25 genes. Prior to selection of novel XLMR-candidate genes from these 25 genes, X-inactivation analysis was conducted. It has been shown that up to 50% of XLMR carriers show skewed X-chromosome inactivation, with preferential activation of the 'normal' allele (Plenge et al. 2002). Skewed X-inactivation in the carrier mother (Fx67.2) of this family would support the strategy to select and screen novel candidate MR genes in this region. However, X-inactivation analysis in this female showed an X-inactivation ratio of 72:28, towards selective inactivation of the 'disease-associated' allele. It has been demonstrated that this mild degree of skewing is present in female control subjects (Plenge et al. 2002). Plenge and colleagues (2002) demonstrated that X-inactivation ratios in excess of 70:30 were present in 30 female control subjects as compared to 63 XLMR carrier females (Plenge et al. 2002). This makes XLMR carrier females approximately twice as likely to have skewed X-inactivation as compared to the background population. Therefore, the 72:28 X-inactivation ratio was suggestive of the disorder being X-linked in family Fx67. However, given the genetic heterogeneity and diverse functions in which XLMR gene products participate it was exceptionally difficult to prioritise XLMR-candidate genes. Finally, the LOD score in this family was 0.60 (θ=zero), a figure which was not statistically significant, but merely suggestive of linkage to this region given the co-segregation of alleles in the affected males. Taken together, it was decided that it was not feasible to screen novel XLMR-candidate genes in the region at this time. It is envisaged that as next generation DNA sequencing technologies become more affordable, screening of candidate-XLMR genes in this region will be possible.

In conclusion the positional candidate gene approach in this study has led to the identification of an *ATRX* disease-causing mutation in family XMR2. Also, with the establishment of the *ATRX* 'hotspot' mutation detection protocol, testing can now be offered in the diagnostic setting through the NHLS (National Health Laboratory Services) for clinically suitable MR patients. The enhancement of the repertoire of diagnostic protocols offered to MR patients will lead to better genetic management and counselling for those families afflicted with MR. Finally, given the proposed role of ATRX as a chromatin remodeler as well as the family size, XMR2 provided an excellent model for whole-genome DNA methylation investigations (Chapter 6).

CHAPTER 5

FUNCTIONAL CANDIDATE GENE MUTATION DETECTION



5.1 INTRODUCTION

The 21st century has seen a rapid expansion in the molecular tools and technologies available to molecular geneticists, concomitantly there has been an explosion in the number of genetic disorders which can now be attributed to a molecular cause. XLMR disorders have been no exception, in the last 5 years 42 genes have been found to cause XLMR, a figure comparable to the 45 XLMR genes identified since the first XLMR gene, *FMR1*, was identified in 1991 (i.e. 1991-2004) (Chiurazzi, Tabolacci and Neri. 2004, Chiurazzi et al. 2008, Verkerk et al. 1991). Along with enhancing the gene discovery process, these technological advancements have also assisted XLMR gene screening in large cohorts of clinically relevant patients. Such studies have enabled researchers to estimate the mutation frequencies of most of the 87 known XLMR genes, with certain genes being more frequently mutated than others (Chiurazzi et al. 2008, de Brouwer et al. 2007).

FMR1 was the first gene to be associated with XLMR; in 1991 it was shown that a *FMR1* 5'UTR CGG expansion was the molecular cause of the FXS (Verkerk et al. 1991). Since then, the FXS has been established as the most common type of XLMR accounting for 15-25% of all XLMR (Kleefstra and Hamel. 2005). As described previously, of the approximately 450 families referred to the Division of Human Genetics, UCT for FXS testing only 23% of patients have been shown to be positive for the *FMR1* 5'UTR expansion mutation.

The second most commonly implicated XLMR gene is the *Aristaless Related Homeobox* (*ARX)* gene, with reported mutation frequencies of between 6.6 and 9.5% (de Brouwer et al. 2007, Mandel and Chelly. 2004, Poirier et al. 2004). *ARX* mutations account for a diverse range of phenotypes, including non-syndromic MR. Given this relatively high mutation frequency and the clinical heterogeneity associated with *ARX* mutations, molecular *ARX* investigations were conducted in all XLMR families as well as all sporadic MR males who's DNA was available in the Division of Human Genetics, UCT DNA bank.

Internationally, the remaining explained XLMR cases are accounted for by smaller mutation frequencies in a number of genes, including (most commonly): *MECP2*, *OPHN1*, *PQBP1* and *KDM5C* (*JARID1C*) (de Brouwer et al. 2007). Of these genes, *MECP2* duplications were investigated in the UCT MR DNA bank in a separate study. *OPHN1* and *PQBP1* are associated with very specific syndromic forms of MR which were not common in this group of South African patients and were therefore not investigated at this stage (Bergmann et al. 2003, Martinez-Garay et al. 2007). Finally, within the MR families of the EuroMRX consortium DNA bank, *KDM5C* was shown to account for the disorder in 4.2% and 4.3% of XLMR families and sib-pairs respectively (de Brouwer et al. 2007). *KDM5C* mutations have been identified in patients with a subset of clinical features which were also prevalent in the UCT DNA bank (Abidi et al. 2008, Adegbola et al. 2008, Jensen et al. 2005, Santos et al. 2006). Therefore, *KDM5C* was selected as a functional candidate gene in this study for investigation in a clinically stratified XLMR cohort.

In this study, two of the most frequently mutated XLMR genes (*ARX* and *KDM5C*) were investigated in a cohort of South African XLMR patients selected from the UCT DNA bank. The purposes of these investigations were two-fold. Firstly, determining the 'common' XLMR gene mutation frequencies in South African patients provides guidance for the feasibility and practicality of expanding the repertoire of XLMR diagnostic tests. Secondly, in those patients in whom putative 'epigenetic' XLMR gene mutations are identified, DNA methylation profiles can be investigated in order to test the central hypothesis of this dissertation (i.e. whether an alteration to the DNA methylation profile is a molecular feature of MR in patients positive for mutations in putative 'epigenetic' XLMR genes.).

5.1.1 FUNCTIONAL CANDIDATE GENE: *ARX*

5.1.1.1 Phenotypes associated with ARX mutations

The identification of MR-causing mutations in the *ARX* gene were first described in 2002 in patients with both MRX and MRXS (Stromme et al. 2002b). Subsequent studies in numerous cohorts have led to the identification of *ARX* mutations in patients with a clinically diverse phenotypic presentation, including:

- Non-syndromic MR with/without epilepsy (Bienvenu et al. 2002, Gronskov et al. 2004, Laperuta et al. 2007, Partington et al. 2004, Poirier et al. 2005, Poirier et al. 2006, Rujirabanjerd et al. 2007, Stepp et al. 2005, Stromme et al. 2002b, Troester, Trachtenberg and Narayanan. 2007)
- Numerous MR syndromes of which epilepsy is a core feature. These include: West syndrome or X-linked infantile spasms syndrome (ISSX) (OMIM **#308350**), early infantile epileptic encephalopathy-1 (EIEE1), Ohtahara syndrome (OMIM **#308350**), X-linked myoclonic epilepsy with generalized spasticity and intellectual disability (XLMESID) (OMIM **#300432**), as well as ISSX with severe dyskinetic quadriparesis (Guerrini et al. 2007, Kato et al. 2003, Kato et al. 2007, Scheffer et al. 2002, Stromme et al. 2002a, Stromme et al. 2002b, Wohlrab et al. 2005)
- Partington syndrome (OMIM **#309510**) (Frints et al. 2002, Gronskov et al. 2004, Stromme et al. 2002b, Turner et al. 2002)
- X-Linked lissencephaly with ambiguous genitalia (XLAG) (OMIM **#309215**) (Bhat et al. 2005, Kato et al. 2004, Kitamura et al. 2002, Uyanik et al. 2003)
- Gross brain malformations including agenesis of the corpus callosum (ACC) with abnormal genitalia (Proud syndrome) (OMIM **#300004**), hydraencephaly, transsphenoidal encephalocele and hypopituitarism (Kato et al. 2004, Van Esch et al. 2004).

5.1.1.2 ARX: The gene and its protein

The *ARX* gene (OMIM ***300382**) was identified by Stromme et al. (2002) using transcription mapping in a candidate region originally refined by linkage analysis to Xp22 (Stromme et al. 2002b). The *ARX* gene spans a chromosomal region of 12.5kb and consists of five exons. The GC content of this gene is high, reaching 78% in exon 2. Other features include GCG and GCC repetitive regions in exons 2 and 4 respectively. The *ARX* coding DNA sequence is highly conserved amongst various organisms (Bienvenu et al. 2002, Ohira et al. 2002)

The *ARX* gene is flanked at both 3'and 5'ends by stretches of highly conserved nucleotide sequences. These conserved regions are known as ultraconserved elements (UCEs) defined as DNA segments longer than 200bp that are 100% conserved between orthologous sequences in other species (human, rat, mouse). There are 11 UCEs that flank *ARX* and it has been postulated that these elements act as *ARX* transcription enhancers (Bejerano et al. 2004). More recently it has been shown that one of these 11 UCE elements (uc.467), does indeed act as an *ARX* enhancer (Colasante et al. 2008). Collectively this evidence suggests that these 11 UCEs are good candidates for mutation detection in patients with XLMR. However, it should be noted that deletion of one of these UCEs was shown to lead to no recognizable phenotype in knock-out mice (Ahituv et. al. 2007), suggesting that the role of UCEs may be more complex than previously anticipated.

The *ARX* gene transcribes a range of isoforms (Table 5.1). Recently it has been suggested that the discrepancy in transcript sizes is not due to alternative splicing, but rather due to alternate polyadenylation site usage, thus resulting in variable 3'UTR lengths (Gecz, Cloosterman and Partington. 2006). Notwithstanding this size discrepancy, it has been established that expression levels in foetal brain tissue are far higher than the adult counterpart, suggesting an involvement in neuronal maturation and development (Ohira et al. 2002, Poirier et al. 2004).

Table 5.1 Localisation and size of mRNA transcript of the *ARX* gene

	Size of isoforms of the *ARX* mRNA transcript (kb)		
Location	(Stromme et al. 2002b)	(Ohira et al. 2002)	(Bienvenu et al. 2002)
Foetal tissue			
Brain	Not Investigated	3.3	~3
Adult tissue			
Brain	2.8	5.9	~3
Skeletal muscle	2.8, 2.5 and 2.1	5.9	
Heart		5.9	
Liver		5.9	

The ARX protein has an open reading frame of 1686bp and consists of 562 amino acids. ARX is a member of the Aristaless-related subset of the paired (Prd) class of homeodomain proteins (Ohira et al. 2002, Stromme et al. 2002b). This family of transcription factors are known to play a role in cerebral development and patterning (Bienvenu et al. 2002). ARX has been shown to act as a bifunctional transcription

regulator with both repressive and activating roles (Fullenkamp and El-Hodiri. 2008, McKenzie et al. 2007, Seufert, Prescott and El-Hodiri. 2005).

ARX consists of a number of functional domains (Figure 5.1):

- **The octapeptide motif:** located at the N-terminus, is a DNA binding site and is known as the Goosecoid Engrailed Homology (GEH) or eh1 within the Engrailed (En) homeoprotein (Poirier et al. 2004). This DNA binding domain acts as a transcription repressor through interaction with the Groucho/transducin enhancer of split (TLE) family of co-repressors (Fullenkamp and El-Hodiri. 2008, McKenzie et al. 2007)
- **Three nuclear localisation signals (NLSs):** located throughout the ARX protein (Gecz, Cloosterman and Partington. 2006)
- **The homeodomain:** a DNA binding domain which has been shown to repress expression of at least one gene to date (*Pax4*) by binding at the promoter sequence. (Collombat et al. 2005). Also it has been shown that IPO13, a mediator of nuclear import, interacts with the ARX homeodomain (Shoubridge et al. 2007)
- **Four polyalanine tracts**: these elements are common components of transcription factors (Lavoie et al. 2003). Expansions of these polyalanine tracts in other transcription factors (e.g. *SOX3*, *HOXD13*, *FOXL2*, *ARX*) are associated with a range of genetic disorders (Albrecht and Mundlos. 2005). While the function of these alanine tracts remain largely elusive it has been suggested that they may act as spacer elements which maintain the protein tertiary structure integrity, and/or facilitate protein-protein interactions and/or DNA binding (as reviewed in (Amiel et al. 2004)).
- **The other repression domain (ORD):** the second transcription repression domain (Fullenkamp and El-Hodiri. 2008, McKenzie et al. 2007). This domain is *Groucho*-independent and the repression activity has been shown to be enhanced by C-Terminal Binding Protein 1 (CtBP1) binding (Fullenkamp and El-Hodiri. 2008)
- **The Aristaless domain:** was shown to act as a transcriptional activator (McKenzie et al. 2007).

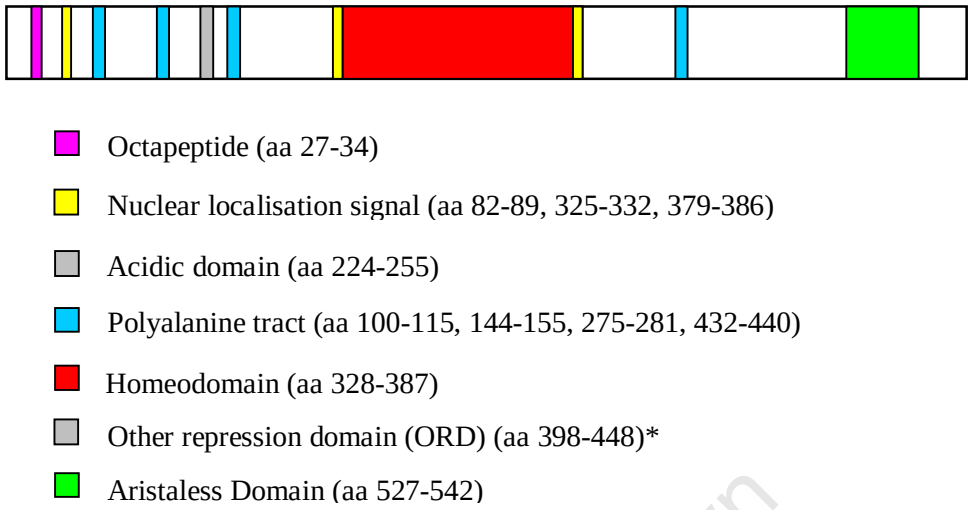


Figure 5.1 A schematic representation of the domains of the ARX protein (Gecz, Cloosterman and Partington. 2006).*as reported in Fullenkamp and El-Hodiri. 2008.aa=amino acid

5.1.1.3 X mutations associated with XLMR

To date 34 *ARX* mutations have been identified in 88 families (Appendix 5A). There appears to be a strong genotype-phenotype correlation between the severity of mutations and the corresponding clinical manifestations. This allows mutations to be categorised into two groups, delineated according to the presence or absence of malformation features (clinical details in Appendix 5A):

- Severe mutations are those resulting in a loss of function of the protein. These include mutations leading to protein truncation or non-conservative missense mutations in the homeodomain or Aristaless domain (Figure 5.2). Severe mutations are associated with malformation phenotypes including XLAG and gross malformations of the brain.

- Less severe (mild) mutations are those that have less impact on the functional deficit of the protein. These include conservative mutations in the homeodomain, mutations outside the homeodomain and Aristaless domain, C-terminal truncations encompassing the Aristaless domain, as well as expansions of the polyalanine tracts (Figure 5.2). Mild mutations are associated with NS-XLMR as well syndromic manifestations including West syndrome, Partington Syndrome and XLMESID.

While there is evidently a clear genotype–phenotype correlation for *ARX* mutations, there is one exception described to date. Van Esch and colleagues (2004) identified an individual with severe brain malformations (Transsphenoidal encephalocele and hypopituitarism) who carries the normally 'mild' *ARX* mutation, c.428_451dup24 (Van Esch et al. 2004).

It is also interesting to note that all expansion mutations in the first polyalanine tract (c.304ins$(GCG)_7$ and c.298_330dupGCGGCA$(GCG)_9$) are associated with MR syndromes of which epilepsy is a core feature e.g. ISSX, EIEE and West syndrome (Appendix 5A). Several smaller expansions of this tract have also been reported in patients with NS-XLMR with absence of seizures (c.304ins$(GCG)_1$, c.304ins$(GCG)_2$ and c.304ins$(GCG)_3$) (Bienvenu et al. 2002, Gronskov et al. 2004). These findings led to the suggestion that the longer the expansion the more severe the clinical presentation and the earlier the onset of the disorder. However, the c.304ins$(GCG)_1$ expansion was identified in a normal male control DNA sample and also did not segregate with the disorder in one family (Gronskov et al. 2004). The c.304ins$(GCG)_1$ mutation is therefore not a disease-causing mutation. To date, the larger polyalanine expansions have not been identified in normal controls, suggesting a threshold effect for the number of alanines tolerated.

Expansions of the first polyalanine tract account for 14 out of the 88 (16%) known *ARX* mutations. However, the most common *ARX* mutation is the expansion from 12 to 20 alanine residues in the second polyalanine tract (c.428_451dup24), accounting for 40/88 of cases (45%). This duplication exhibits both inter- and intra- familial clinical heterogeneity and has been detected in a number of clinical presentations including: MRX, Partington syndrome, ISSX and West syndrome (Appendix 5A).

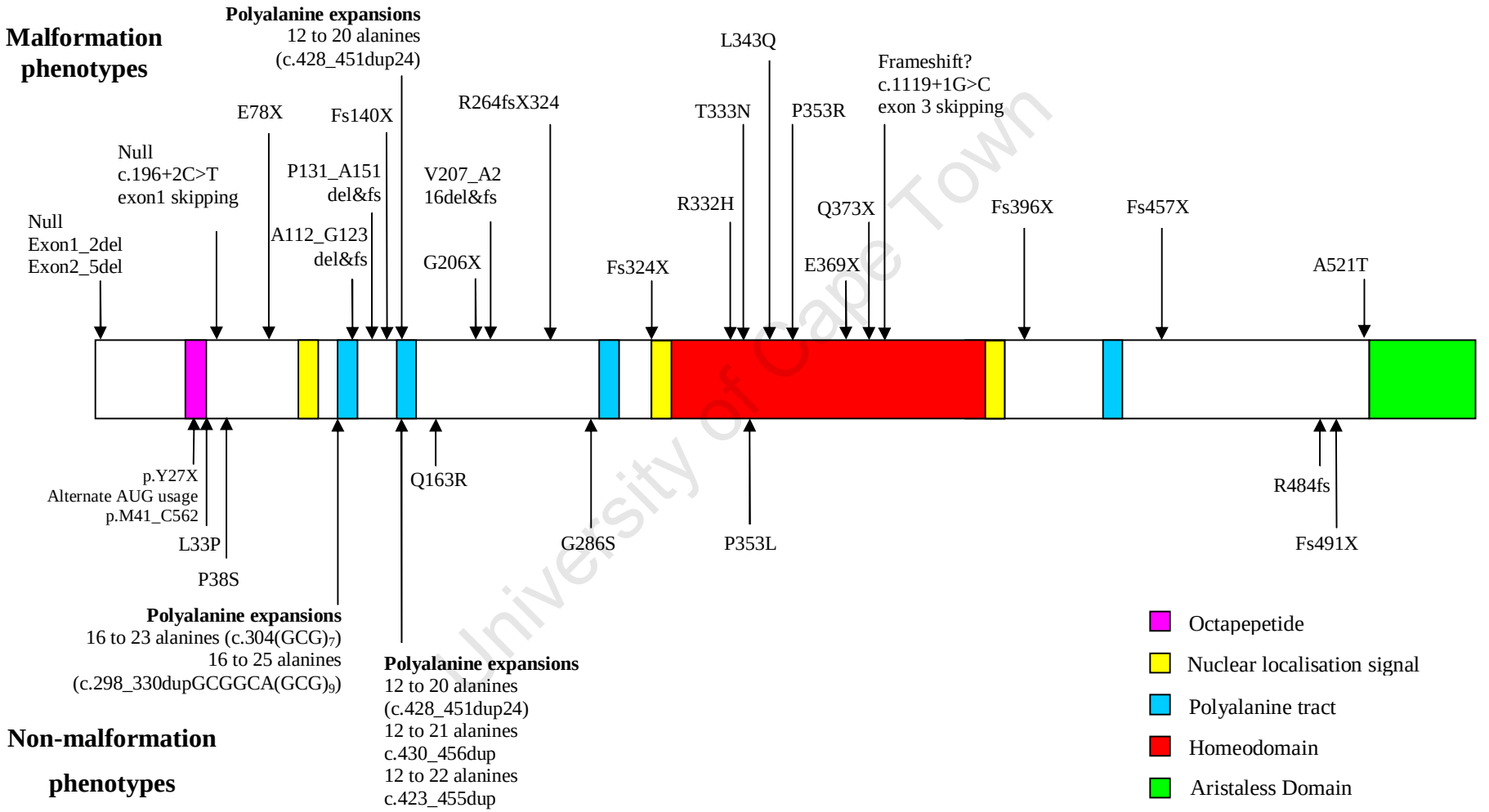


Figure 5.2 Schematic representation of known mutations in the *ARX* gene. Malformation phenotypes include XLAG as well as other gross malformations of the brain. Mild mutations are those responsible for non-malformation phenotypes including MRX, Partington Syndrome, West syndrome and XMESID

5.1.1.4 The pathophysiological basis of ARX mutations

Several lines of evidence support the role of ARX as a bi-functional transcription repressor and activator (Fullenkamp and El-Hodiri. 2008, McKenzie et al. 2007, Seufert, Prescott and El-Hodiri. 2005). In support of this function, ARX mutant constructs (corresponding to those seen in XLMR patients) in the two repressive domains (the octapeptide and the ORD) exhibited reduced repressive activity (Fullenkamp and El-Hodiri. 2008, McKenzie et al. 2007). In contrast, ARX mutants encompassing the Aristaless domain showed up to a 14 fold increase in reporter activity, supporting the hypothesis that this domain acts as a transcription activator (McKenzie et al. 2007). These results show that *ARX* mutations which occur within these transcriptional regulatory elements disrupt the bi-functional regulator role of ARX, resulting ultimately in the MR phenotype.

A significant percentage (61%) of *ARX* mutations can be attributed to expansion of either one of the first two polyalanine tracts. The expansion of polyalanine tracts in other proteins (particularly transcription factors) beyond a certain threshold leads to protein misfolding, aggregation and degradation, ultimately resulting in cell death. The pathological consequences of this protein aggregation is thought to be the molecular mechanism underlying a number of genetic disorders (as reviewed in Albrecht and Mundlos. 2005). There are however, conflicting reports as to whether the polyalanine expansions in ARX result in protein aggregation. The first investigation for polyalanine nuclear inclusions found that expansions in the first polyalanine tract (c.304ins$(GCG)_7$) did not lead to intranuclear inclusions (Poirier et al. 2004). However, subsequently, two groups demonstrated that c.304ins$(GCG)_7$ forms aggregates in the cytoplasm (Gecz, Cloosterman and Partington. 2006, Shoubridge et al. 2007) while another reported intranuclear inclusions (Nasrallah, Minarcik and Golden. 2004). Concerning the expansion of the second polyalanine tract (c.428_451dup), two conflicting studies have been published by the same group with the first reporting polynuclear inclusions (Gecz, Cloosterman and Partington. 2006) and the second showing no aggregation in either the nucleus or cytoplasm (Shoubridge et al. 2007). This discrepancy could be attributed to the different methodologies adopted for each study. Ultimately, due to these conflicting reports

the role of protein aggregation as a disease mechanism in patients with ARX polyalanine expansions remains inconclusive at this stage.

In studies involving *Arx* mutant mice Kitamura et al. (2002) showed that Arx plays a critical role in the development of the forebrain. These Arx mutants showed aberrant GABAergic interneuronal differentiation and impaired migration from the ganglionic eminence to the developing cortex (Kitamura et al. 2002, Kitamura et al. 2009). Subsequent studies in additional mutant/ knock-out *Arx* mice and *xArx* frogs morpholinos have corroborated the critical role of ARX in the development of the human forebrain, particularly in GABAergic and cholinergic neuronal migration and basal ganglia differentiation (Colombo et al. 2004, Colombo et al. 2007, Seufert, Prescott and El-Hodiri. 2005).

Furthermore, transcriptome analysis from *Arx* mutant mice by Fulp et al. (2008) demonstrated dysregulation of 84 genes in the absence of Arx. Interestingly, this dysregulated data set was enriched for genes involved in biological processes involving neuronal migration, neurogenesis, transcription regulation and axonal guidance. Also, a number of the human orthologues of the identified *Arx* transcription targets have been associated with neurological disorders such as epilepsy, autism and MR. This study further emphasises the role of Arx in neuronal development and provides interesting candidate genes for mutation screening in patients with MR (Fulp et al. 2008).

5.1.2 FUNCTIONAL CANDIDATE GENE: *KDM5C* (*JARID1C*)

5.1.2.1 Phenotypes associated with KDM5C mutations

In keeping with the diverse clinical spectrum of mutations in XLMR genes, the phenotypes associated with *KDM5C* mutations are broad, encompassing both MRX and MRXS. The severity of MR associated with *KDM5C* mutations ranges from mild to severe and includes the manifestation of a number of additional clinical features (Table 5.2). From the six studies published to date there appears to be a preponderance of certain clinical features in *KDM5C* mutation positive patients, these include speech impairment (45%), short stature (40%), hyperreflexia or

spasticity (35%), aggressive behaviour (30%), epilepsy/seizures (30%) and hand abnormalities (30%) (Abidi et al. 2008, Adegbola et al. 2008, Jensen et al. 2005, Rujirabanjerd et al. 2009, Santos et al. 2006, Tzschach et al. 2006).

Table 5.2 A summary of the clinical features described in the 20 *KDM5C* mutation positive families reported to date (Abidi et al. 2008, Adegbola et al. 2008, Jensen et al. 2005, Rujirabanjerd et al. 2009, Santos et al. 2006, Tzschach et al. 2006).

Clinical feature	**Familial incidence (%)**
MR	20/20 (100)
Speech impairment	9/20 (45)
Short stature	8/20 (40)
Hyperrreflexia/ spasticity	7/20 (35)
Aggression	6/20 (30)
Epilepsy/ seizures	6/20 (30)
Hand abnormalities (Broad hand/tapering fingers, Camptodactly and clinodactyly, Large fingers/proximal thumb, brachydactyly)	6/20 (30)
Strabismus	4/20 (20)
Microcephaly	4/20 (20)
Prominent ears	3/20 (15)
Small testes	2/20 (10)
High narrow plate	2/20 (10)
Cafe au Lait spot/ abnormal skin pigmentation	2/20 (10)
Macrocephaly	1/20 (5)
Facial hypotonia	1/20 (5)
Overfriendly/anxious	1/20 (5)
Autism spectrum disorder	1/20 (5)
Club feet	1/20 (5)
High prominent nasal bridge	1/20 (5)
Flexion contractor	1/20 (5)

5.1.2.2 *KDM5C*: *The gene and its protein*

The association of *lysine (K)-specific demethylase 5C* (*KDM5C*) (alias *SMCX*, *JARID1C*) with XLMR was first identified in 2005 by systematic mutation screening of 47 candidate genes in XLMR families with overlapping Xp11 linkage intervals (Jensen et al. 2005). Located at Xp11.22, *KDM5C* consists of 26 exons and spans a length of 34kb.

Investigations pertaining to the X-inactivation status of *KDM5C* in females have produced conflicting results. Several studies have shown *KDM5C* escapes X–inactivation in both mouse and the human genomes (Agulnik et al. 1994, Carrel and Willard. 2005, Li and Carrel. 2008, Murakami et al. 2009). However, another study showed partial X-inactivation at different developmental stages and in certain adult tissues (Sheardown et al. 1996). Finally, females who are carriers of disease-causing *KDM5C* mutations exhibit a skewed X-inactivation pattern suggesting the gene is subject to X-inactivation (Abidi et al. 2008).

KDM5C encodes for a transcript of ~6kb and is expressed in a variety of tissues. Expression was highest in the brain and skeletal muscle, with moderate expression in the lung and minimal expression in the pancreas, heart and liver (Jensen et al. 2005).

The *KDM5C* transcript consists of an open reading frame of 4680bp and translates into a 1560 amino acid protein. KDM5C forms part of the JARID1 (Jumonji (Jmj) AT-rich interactive domain) family. This protein has high homology to three other JARID1 protein family members, JARID1A/B/D with 51%, 47% and 85% amino acid conservation respectively (Jensen et al. 2005).

KDM5C is a transcription regulator and chromatin remodeler and has recently been shown to act as a histone demethylase responsible for the demethylation of di/trimethylated- histone H3 lysine 4 ($H3K4_{me2}$ and $H3K4_{me3}$) (Iwase et al. 2007, Tahiliani et al. 2007). The protein consists of a number of functional domains including (Figure 5.3.):

1. **The JmjN domain:** common to a subset of the JmjC protein family and maintains the structural integrity of the protein, particularly the JmjC domain, and hence promotes optimal catalytic activity of *KDM5C* (Chen et al. 2006)
2. **The AT-rich domain-interacting domain (ARID):** the function of this domain within KDM5C is largely unknown. However, ARID has been shown to bind DNA in a sequence-specific and non-sequence-specific manner suggesting a role for KDM5C in DNA binding (Kortschak, Tucker and Saint. 2000)
3. **Two Plant Homeodomain (PHD) zinc fingers:** (designated PHD1 and PHD2). PHD1 motif tethers KDM5C to tri-methylated histone H3 lysine 9 (H3K9me3) (Iwase et al. 2007). The function of the C-terminal PHD2 remains elusive
4. **The JmjC domain:** catalyses the demethylation of $H3K4_{me3}$ and $H3K4_{me2}$ (Iwase et al. 2007)
5. **The C5HC2 zinc finger:** the function of this domain remains unclear but DNA binding is a putative function.

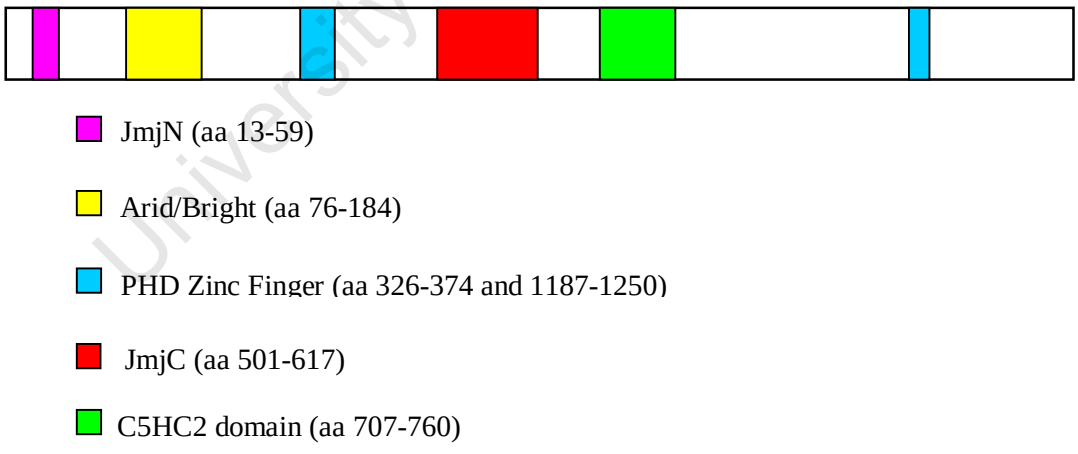


Figure 5.3 A schematic representation of the domains of the KDM5C protein (Tzschach et al 2006)

KDM5C is a histone demethylase responsible for demethylation of $H3K4_{me3}$ and $H3K4_{me2}$ through the catalytic activity of the JmjC domain. This demethylase activity is directed by the binding of the PHD1 domain to H3K9me3 (Iwase et al. 2007). Trimethylation of H3K4 and H3K9 have opposite effects on transcription regulation, with H3K4me3/me2 being associated with gene activation and H3K9me3 with a repressive state. Therefore, KDM5C facilitates the cross-talk which occurs between different histone residues in order to co-ordinate a particular expressive state where protein binding to $H3K9_{me3}$ (repressive) leads to demethylation of H3K4me3 to $H3K4_{me2}$ and $H3K4_{me1,}$ hence inducing a repressive chromatin state (Figure 5.4).

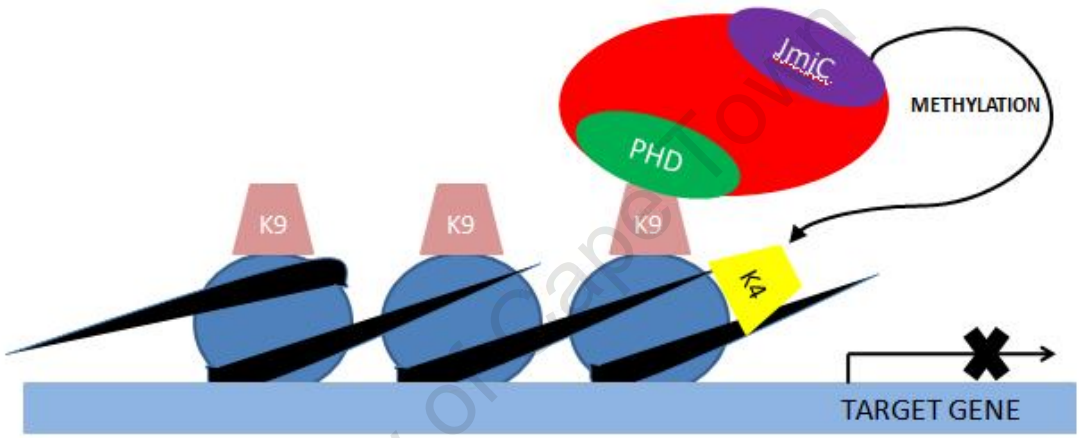


Figure 5.4 The histone H3K4 demthylase activity of KDM5C. By binding of the KDM5C PHD domain to the repressive H3K9me3 epigenetic mark the JmjC domain catalyses the demethylation of H3K4me3 and H3K4me2 to further induce a repressive chromatin state and prevent target gene expression.

To further support the role of KDM5C as a chromatin remodeler, chromatin immunoprecipitation (ChIP) experiments from nuclear extracts show that the protein exists in complex with numerous other chromatin modifiers, including RE-1 Transcription Factor (REST) (Tahiliani et al. 2007). REST is responsible for the repression of neuronal genes in non-neuronal cell types. Using ChIP assays it was shown that KDM5C and REST were both present at RE-1 target sites, providing further evidence that these two proteins interact to control the expression of neuronal genes (Tahiliani et al. 2007).

5.1.2.3 KDM5C mutations associated with XLMR

To date 20 families comprising a total of 44 affected males and four carrier females with mild intellectual disability have been reported with mutations in *KDM5C* (Appendix 5B) (Figure 5.5) (Abidi et al. 2008, Adegbola et al. 2008, Jensen et al. 2005, Rujirabanjerd et al. 2009, Santos et al. 2006). Of the 282 families screened for *KDM5C* mutations in the European XLMR consortium, a total of 13 families carry mutations in *KDM5C* (de Brouwer et al. 2007, Rujirabanjerd et al. 2009). This equates to a frequency of 4.6% in this population group, resulting in *KDM5C* being quoted as one of the genes more commonly implicated in XLMR.

Mutations in *KDM5C* do not appear to present with a clear genotype-phenotype correlation. All truncating mutants identified to date result in severe MR. However, missense mutations in the functional domains result in both severe and mild MR. The majority of mutations cluster between amino acids 332 and 766 encompassing the three functional domains, PHD1, JmjC and C5HC2 (which correlates to exons 4-16). Fourteen of the 20 (70%) mutations identified to date cluster in this region. Future investigations in larger cohorts will determine whether this region constitutes a true 'mutation hotspot' as well as shed light on any potential genotype-phenotype correlation.

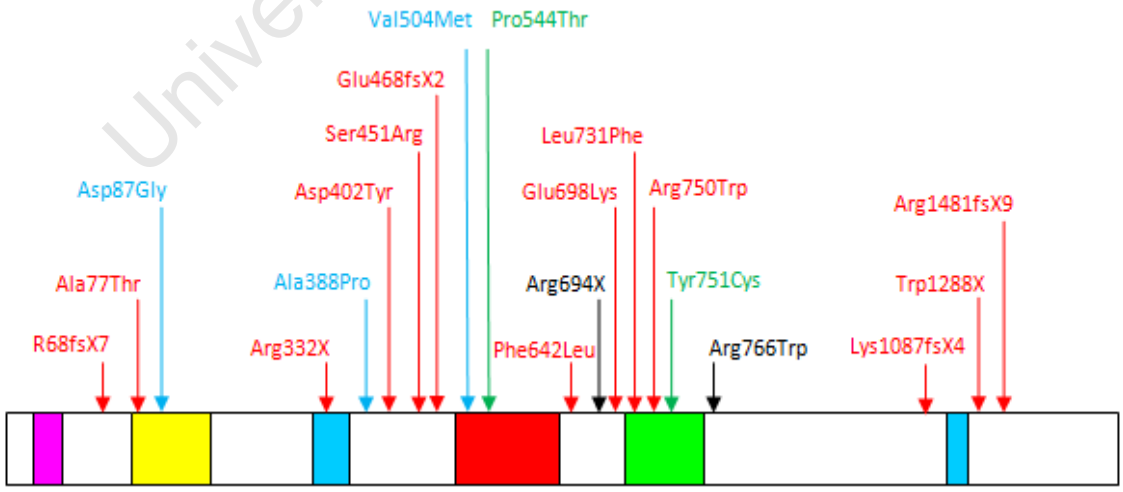


Figure 5.5 Schematic of disease-associated mutations in *KDM5C*. The functional domains are represented as in Figure 5.4. Colour of text indicates the severity of MR: red – severe, green – moderate, blue – mild, black- not reported.

5.1.2.4 The pathophysiological basis of KDM5C mutations

KDM5C is a histone demethylase responsible for the demethylation of $H3K4_{me3}$ and $H3K4_{me2}$ and exists in a protein-complex (including REST) that represses neuronal gene expression in non-neuronal tissues (Iwase et al. 2007, Tahiliani et al. 2007). Iwase et al. generated four mutant KDM5C recombinant proteins which represented four known XLMR-causing mutations (Asp87Gly, Asp402Tyr, Glu698Lys and Tyr751Cys). H3K4 demethylase activity was impaired in these mutants in a manner that seemed to be directly proportional to the severity of MR (Tahiliani et al. 2007). These results suggest that the pathogenesis of XLMR in *KDM5C* mutation positive patients is due to ectopic neuronal gene regulation.

Animal models for *Kdm5c* mutants have been generated in order to better understand the pathophysiological effects of gene mutations. *Kdm5c* knockdown in the zebrafish resulted in neurodevelopmental defects including aberrant brain patterning and autonomous neuronal cell death (Iwase et al. 2007). In rats, conditional *KDM5C* knockdown (restricted to the primary granules) did not result in neuronal cell death, but rather a significant decrease in dendritic length (Iwase et al. 2007). This dendritic morphological abnormality is also commonly seen in patients with XLMR (Ropers and Hamel. 2005). These studies show that KDM5C plays a critical role in neuronal development and that mutations in this gene lead to MR, possibly through misregulation of neuronal gene expression.

5.2 METHODS

5.2.1 FUNCTIONAL CANDIDATE GENE: *ARX*

5.2.1.1 ARX mutation detection cohort selection

Owing to the clinical heterogeneity associated with *ARX* mutations as well as their purported high frequency, clinical presentation of patients (besides MR) was not a criterion for *ARX* mutation detection selection. Instead cohorts were stratified according to the MR inheritance pattern which in turn determined the adopted *ARX* screening protocol. The first two cohorts consisted of 77 affected males from families showing a clear or putative X-Linked inheritance pattern and 36 affected males that formed part of a sib-pair. These cohorts were subject to mutation detection across the entire *ARX* coding region (Appendix 5C). A third cohort of 183 patient samples, each representing isolated cases of MR, was screened for the common *ARX* polyalanine expansions (c.428_451dup c.304ins$(GCG)_7$) (Appendix 5C). All patients were previously shown to be mutation negative for the *FMR1* 5'UTR CGG expansion resulting in Fragile X syndrome.

5.2.1.2 PCR amplification of the five ARX exons

Six sets of PCR primers were used to amplify the five *ARX* exons (exon 2 consisted of two overlapping fragments due to the exon's large size) (Appendix 5D). Significant difficulty was experienced in *ARX* exon amplification (particularly exon 2) owing to the gene's high GC content. For these reasons PCR amplification of each exon required different PCR reagents and amplification profiles (Appendix 5E).

5.2.1.3 dHPLC analysis of the five ARX exons

dHPLC analysis was conducted as per manufacturer's instructions [Transgenomic®]. Optimal chromatographic methods were designed using WAVEMAKER™ software [Transgenomic®]. The selected method temperatures for each individual exon were as follows: exon 1: 62.4ºC and 65.4ºC, exon 2a: 69ºC, 68.2ºC and 71.6ºC, exon 2b: 65.4ºC and 66.6ºC, exon 3: 62ºC and 63.4ºC, exon 4: 63.5ºC, 66.9ºC and 68.1ºC and exon 5: 66.8ºC.

Patient-amplified DNA (test) PCR samples were mixed in a 1:1 ratio with a wild type sample, denatured at 95ºC for 5 minutes on a Hybaid touchdown thermal cycler and allowed to re-anneal on the block overnight. A total volume of 5-7µl of the re-annealed heterogeneous solution was injected onto the column. Analysis of dHPLC profiles was conducted using WAVEMAKER™ software [Transgenomic®].

5.2.1.4. DNA sequencing of ARX dHPLC variants

All samples exhibiting deviant dHPLC profiles were subject to Sanger DNA sequencing in order to elucidate the underlying DNA sequence variation. DNA sequencing of the sample exhibiting a deviant profile in *ARX* exon 3 was performed as described previously (Section 4.2.1.4). However, DNA sequencing of variants with altered dHPLC profiles in exon 2A required extensive optimisation and adjustments to the sequencing protocol. Ultimately the optimised DNA sequencing reaction consisted of: 40µM of a nested reverse primer (sequence supplied by Prof J. Gecz, 5'-ctcggtgccggtgccaccac-3) or 40µM of the aforementioned forward primer, 500ng of purified PCR product, 1× Big Dye terminator sequencing buffer [Applied Biosystems], 8µl Big Dye terminator mix [Applied Biosystems] and 10% glycerol [Merck]. Sequencing reactions were cycled through an initial denaturation at 98ºC (10min), followed by 25 cycles of denaturation at 98ºC for 1min, primer annealing at 55ºC for 1min and elongation at 72ºC for 2min, on a GeneAmp® PCR system 9700 thermal cycler [Applied Biosystems].

5.2.1.5 c.428_451dup and c.304ins(GCG)$_7$ PCR-based screen

The same PCR amplification protocol as that of exon 2a was used for amplification of the c.428_451dup and c.304ins(GCG)$_7$ mutations with the following exceptions:

- 0.1 units of GoTaq polymerase [Promega] was used instead of ELT
- The forward primer: 5'gctcccctaagagcaggagg 3' (so as to amplify a smaller fragment which was easier to separate by gel electrophoresis) was used

PCR products were analysed on a 0.8% agarose gel and visualised with EtBr. A c.428_451dup mutation negative and positive control samples were used, these samples produced bands of 477bp and 501bp respectively. No positive control was

available for the c.304ins$(GCG)_7$ mutation, but the expected product size of 498bp was expected to be distinguishable from the wildtype (477bp product), although not from the c.428_451dup (501bp product).

5.2.2 FUNCTIONAL CANDIDATE GENE: *KDM5C*

5.2.2.1 KDM5C mutation detection cohort selection

A cohort of 25 patients (Appendix 5C) was selected for the *KDM5C* mutation screen, the criteria by which these individuals were selected is given below:

- A Clear X-linked recessive inheritance of mental handicap
- Negative for the Fragile X syndrome expansion mutation
- Negative for mutations in the *ARX* gene
- A clinical presentation indicative of *KDM5C* mutations (Table 5.2).

5.2.2.2 PCR amplification of the 26 KDM5C exons

Primers were designed in order to amplify all 26 exons, as well as the intron-exon boundaries of the *KDM5C* gene. Due to *KDM5C* gene structure it was often possible to design amplicons encompassing adjacent exons as well as the intervening intron. For these reasons only 20 amplicons were required to amplify the entire coding region of *KDM5C* (Appendix 5D). PCR amplification of all exons was performed in a 25µl reaction volume using 100ng of genomic DNA, 10µM of each primer, 2.5µM dNTP's [Bioline], 1× GoTaq buffer [Promega] and 1U GoTaq Polymerase [Promega]. The PCR profile consisted of an initial denaturation step at 95ºC for 3 min, followed by 30 cycles at 95ºC for 30s, 59ºC for 30s, and 72ºC for 30s. A final elongation at 72ºC for 3 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation]. All PCR products were checked by electrophoresis on a 1.5% agarose gel and visualised with EtBr.

5.2.2.3 dHPLC analysis of the 26 KDM5C exons

Mutation detection in all *KDM5C* amplicons by dHPLC was performed as described previously for the *ARX* exons (Section 5.2.1.3). The selected method temperatures for each individual exon appear in Table 5.3.

Table 5.3 Temperatures of the various dHPLC methods employed in the screening of the 26 *KDM5C* exons.

***KDM5C* exons**	**Temperature 1 (**ºC)	**Temperature 2 (**ºC)	**Temperature 3 (**ºC)
1	64.1	64.9	
2	62.3		
3	60.2		
4	59	60.9	
5	**60.1**	**62.5**	
6&7	58.8	60.9	
8	61.3	63	
9&10	59.4	61.4	
11&12	59.7	60.5	
13&14	64	66	
15&16	62		
17	61.6	64.5	
18	63.4		
19	63	64.2	
20	61.5	62.5	
21&22	61.7	62.6	
23	62.5	63.1	65.1
24&25	61	63.1	
26	64	67.5	

5.2.2.4 DNA sequencing of KDM5C dHPLC variants

DNA sequencing was performed as before for *ATRX* amplicons (Section 4.2.1.4).

5.2.2.5 Screening of KDM5C intronic variations in a control population

One hundred X chromosomes were screened for *KDM5C* intronic DNA sequence changes to attain an indication of the variants' allele frequency in the relevant population. Controls were matched for gender and ethnicity. These control screens were performed using restriction enzyme digests and/or amplification refractory mutation system (ARMS)-PCR.

BbrP1 digest to detect the c.351-38C>T variation

PCR amplification of exon 4 in 100 control samples was performed as described previously (Section 5.2.1.2). The c.351-38C>T variation destroys a BbrP1 restriction site within the PCR product of *KDM5C* exon 4. PCR products were digested with BbrP1 [Roche] according to manufacturer's instructions [Roche]. Restriction enzyme digest products were electrophoresed on a 2% agarose gel and visualised with EtBr.

ARMS-PCR to detect the c.2517-7_8insTAC variation

A reverse primer sequence was designed to bind specifically to the wild type sequence at position c.2571-7 of the *KDM5C* transcript (Figure 5.6). The PCR was performed in a final reaction volume of 25µl and consisted of 2.5µM dNTP's [Bioline], 1× GoTaq buffer [Promega] and 1U GoTaq Polymerase [Promega]. Primer concentrations were as follows: *KDM5C* exon18 forward (10µM) and reverse primer (5 µM) and 10µM of the wildtype specific primer (5' ctgtgggggctatgaagtatc 3'). The PCR profile consisted of an initial denaturation step at 95ºC for 3 min, followed by 30 cycles at 95ºC for 30s, 59ºC for 30s, and 72ºC for 30s and a final elongation at 72ºC for 3 min on a Px2 thermal cycler [Thermo Electron Corporation]. PCR products were electrophoresed on a 3% agarose gel and visualised with EtBr.

Dde1 digest to detect the c.2623-45G>A

PCR amplification of exon 19 in 100 control samples was performed as described previously (Section 5.2.1.2). The c.2623-45G>A variation destroys a Dde1 restriction site within the PCR product of *KDM5C* exon 19. PCR products were digested with Dde1 [New England Biolabs] according to manufacturer's instructions [New England Biolabs]. Restriction enzyme products were separated using a 12% polyacrylamide gel electrophoresis. (PAGE) and visualised by silver staining.

ARMS-PCR to detect the c.2623-45G>A

An ARMS-PCR was designed so as to screen for the c.2623G>A variant in the control population as an alternative to Dde1 digestion which was unsuccessful. In a manner similar to that illustrated in Figure 5.6 a reverse primer sequence was designed to bind specifically to the wild type sequence at position c.2623-45 of the *KDM5C* transcript. The PCR was performed in a final reaction volume of 25µl and

consisted of 2.5µM dNTP's [Bioline], 1× Go Taq buffer [Promega] and 1U Go Taq Polymerase [Promega]. Primer concentrations were as follows: 10µM *KDM5C* exon18 forward and reverse primer as well as 10µM of the wildtype specific primer (5' aactcctcagctgggcccac 3'). The PCR profile consisted of an initial denaturation step at 95ºC for 3 min, followed by 30 cycles at 95 ºC for 30s, 58 ºC for 30s, and 72 ºC for 30s. A final elongation at 72 ºC for 3 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation]. PCR products were electrophoresed on a 3% agarose gel and visualised with EtBr.

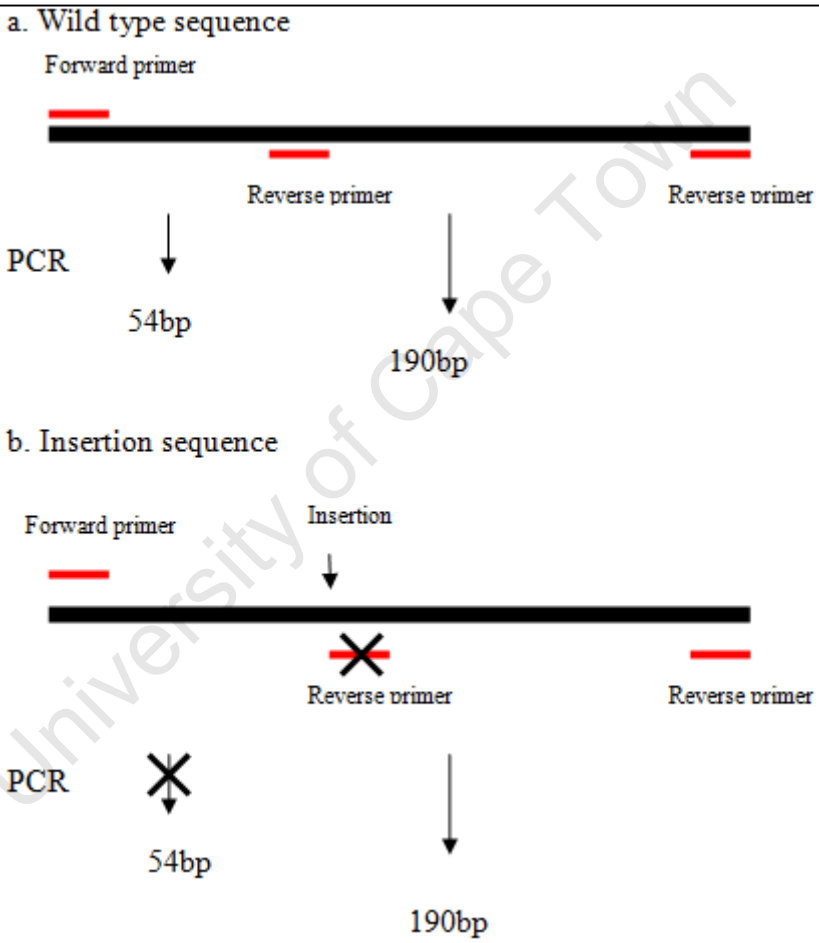


Figure 5.6 ARMS-PCR used to detect the presence of c.2517-7_8insTAC in *KDM5C*. (a) In the presence of the wildtype both reverse primers bind to produce two PCR products of 54 and 190bp. (b) The inhibition of PCR amplification due to restriction of reverse primer binding owing to the TAC insertion sequence, therefore leading only to the single 190bp product.

5.2.2.6 Screening for an intronic insertion in a XLMR patient cohort

Affected males from 76 XLMR families were screened for the c.2517-7_8insTAC variation in order to further elucidate the disease-association of this sequence variant. This was achieved using ARMS-PCR as described above (Section 5.2.1.5.)

5.3 RESULTS

5.3.1 MUTATION DETECTION IN ARX

5.3.1.1 Mutation detection in ARX using dHPLC

Mutation detection in the five *ARX* exons in a total of 113 affected males (77 XLMR families and 36 sib-pairs) revealed no altered dHPLC profiles in exons 1, 2b, 4 and 5. However, three samples (Fx362.1, Fx391.1, Fx446.1) showed deviant dHPLC profiles for exon 2a (Figures 5.7). In addition, an altered dHPLC pattern was detected in exon 3 in sample Fx378.1 (Figure 5.8).

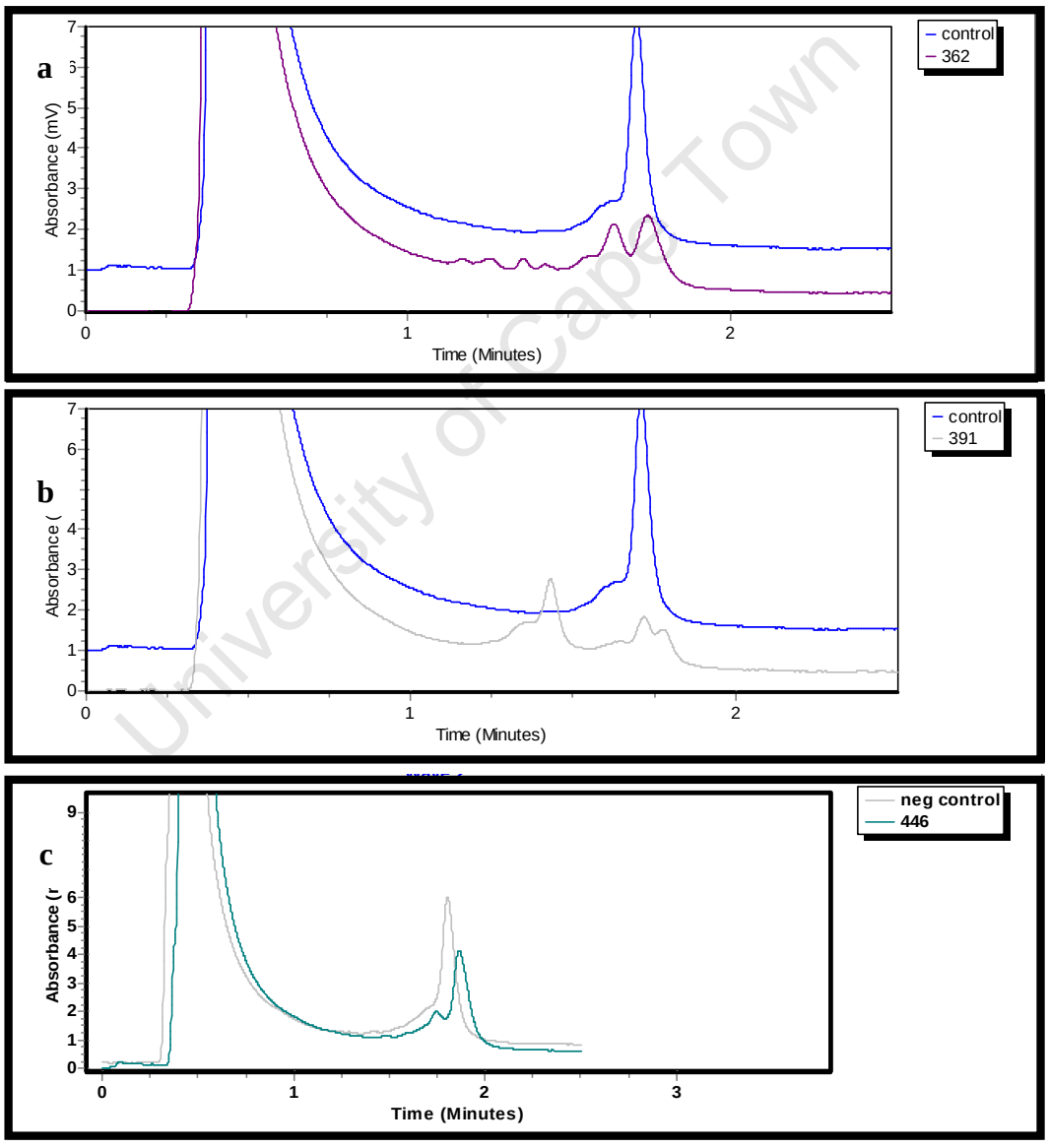


Figure 5.7 Altered dHPLC profiles in *ARX* exon 2a. All PCR products were subject to dHPLC under semi-denaturing conditions, those samples with a variant formed heteroduplexes with the wild-type sample. These heteroduplexes elute at different times due to variable affinity for the column matrix. The resultant deviant profiles at a temperature of 71.6°C are shown for (a) Fx362.1 (b) Fx391.1 (c) Fx446.1.

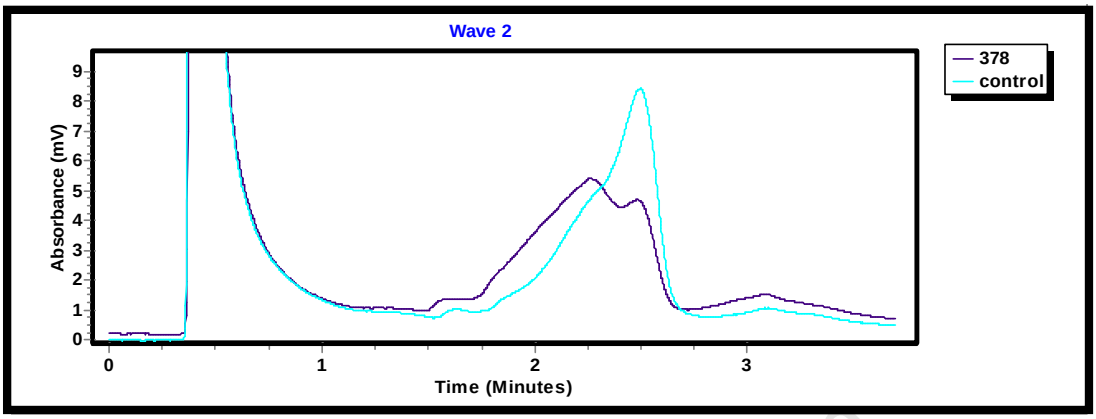


Figure 5.8 Chromatogram of the *ARX* exon 3 variant in sample Fx378.1. PCR products were separated under semi-denaturing conditions. Here, the control (wildtype – blue) sample elutes with only a single peak as compared to Fx378.1 (black) which elutes as two time points hence indicating the presence of a DNA sequence variation Fx378. This dHPLC analysis was performed at 64.3°C.

5.3.1.2 Sequencing and segregation analysis of samples showing aberrant dHPLC profiles

DNA sequencing of those samples exhibiting altered dHPLC profiles revealed the presence of three exonic changes (exon 2a) in samples Fx362.1, Fx391.1 and Fx446.1 and one intronic change (intron 3) in Fx378.1.

<u>The c.300G>A variant in family Fx362</u>

Individual Fx362.1 was shown to carry a c.300G>A transition within the coding region of exon 2a which encodes for a synonymous mutation (Figure 5.9). The c.300G>A was not reported in any of the SNP databases and was therefore a novel nucleotide change. While this variation was a silent mutation, it could not be ruled out that the variant induced aberrant splicing at a putative ESE site. Therefore, the c.300G>A was subject to segregation analysis in the extended family.

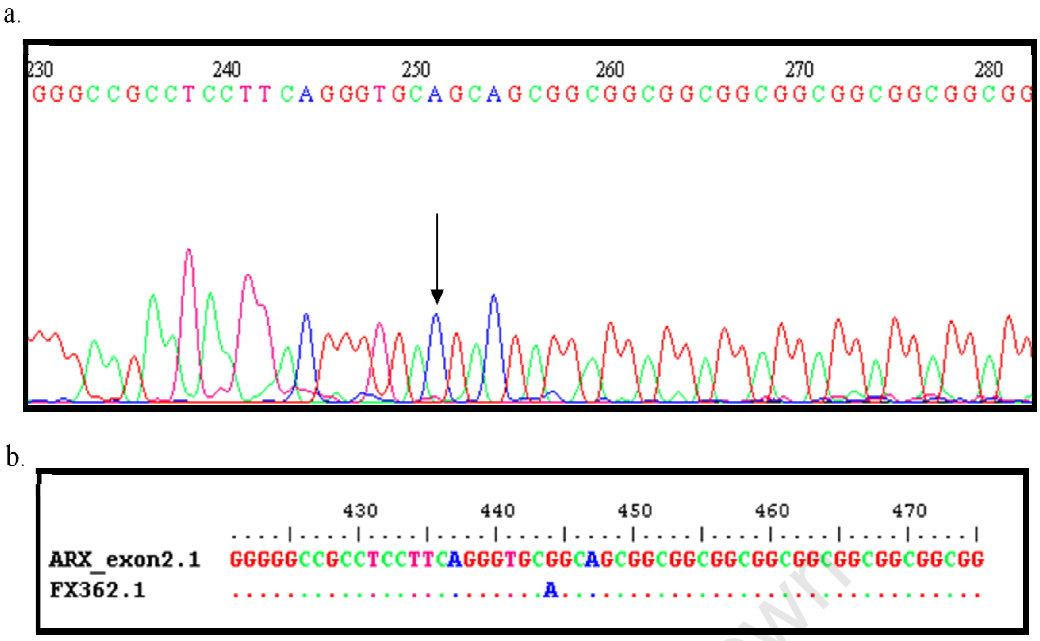


Figure 5.9 DNA Sequence analysis results illustrating the c.300G>A in *ARX* exon 2a in individual FX362.1. (a) An electropherogram of FX362.1, the c.300G>A transition is indicated by the black arrow. (b) An exert of the BioEdit alignment of the consensus sequence (top row) to sample Fx362.1 (bottom row) illustrating the conserved sequence (represented by the dots (.)) as well as the G>A mismatch.

Family Fx362 consisted of three brothers, the two elder brothers (Fx362.1 and Fx362.2) present with NS-XLMR, with MR in the mild-moderate range and macrocephaly. The youngest sib (Fx362.3) exhibited none of the features of his elder brothers and there was no further family history of mental handicap. Segregation analysis of the c.300G>A mutation in *ARX* exon 2a by dHPLC analysis demonstrated that this change did not segregate with the MR phenotype in this family (Figure 5.10). The c.300G>A was originally identified in the patient Fx362.1, but was not detected in either the affected sib (Fx362.2) or the unaffected brother (Fx362.3) by dHPLC analysis. Therefore, the c.300G>A variation was not the disease-causing change in this family.

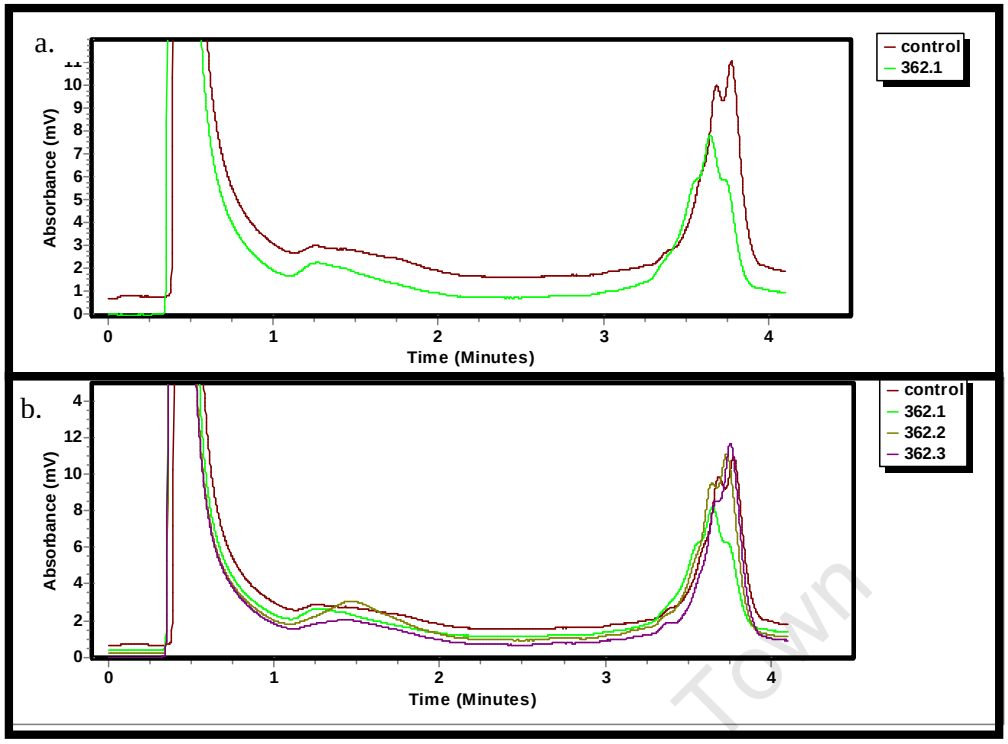


Figure 5.10 Segregation analysis of the c.300G>A *ARX* exon 2a mutation in family Fx362. The chromatogram generated at a temperature of 71.6°C (a) illustrating the aberrant profile of sample FX362.1 (green) as compared to the control (red) indicative of the c.300G>A variation. (b) The chromatogram showing samples Fx362.2 (affected sib) and Fx362.3 (unaffected sib) have the same profile as the control sample and not that of Fx362.1. It was therefore concluded that neither of the sibs of Fx362.1 carry the c.300G>A variation.

The c.428_451dup mutation in families Fx391 and Fx446

Upon DNA sequencing, samples Fx391.1 and Fx446.1 were both shown to be positive for the common *ARX* c.428_451dup mutation (Figure 5.11). It was therefore concluded that the c.428_451dup located in the second polyalanine tract was the cause of XLMR in these two families.

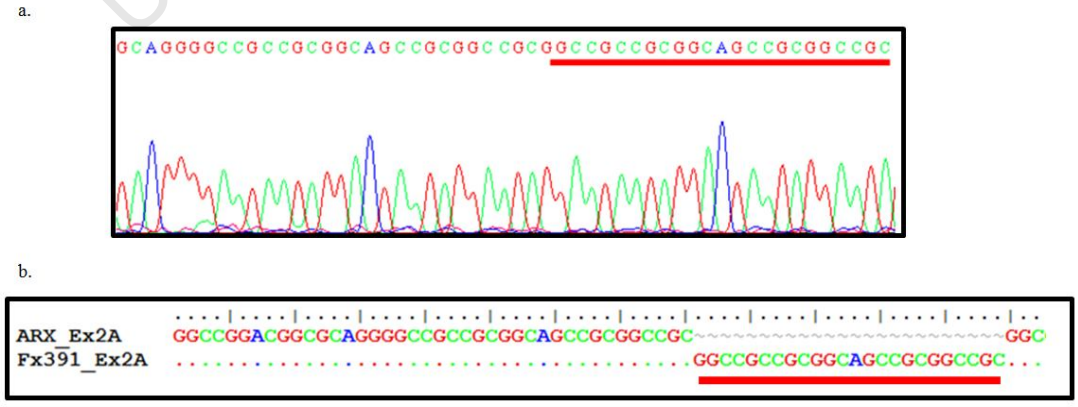


Figure 5.11 Sequence analysis results of *ARX* exon 2a showing the c.428_451dup in individual FX391.1 (a) An electropherogram of sample FX391.1, the c.428_451dup mutation is underlined in red. (b) An exert of the BioEdit alignment of the consensus sequence (top row) to sample Fx391.1 (bottom row) illustrating the conserved sequence (represented by the dots (.)) as well as the c.428_451dup (underlined in red)

Family Fx391 exhibited an extensive history of mental handicap (NS-XLMR in the mild-moderate range) which segregated in an X-linked fashion (Figure 5.12). Carrier testing for the c.428_451dup mutation using dHPLC in the mother (Fx391.2) of the proband (Fx391.1) reflected a positive result (Figure 5.13). No further family members were available for mutation/carrier testing in this family.

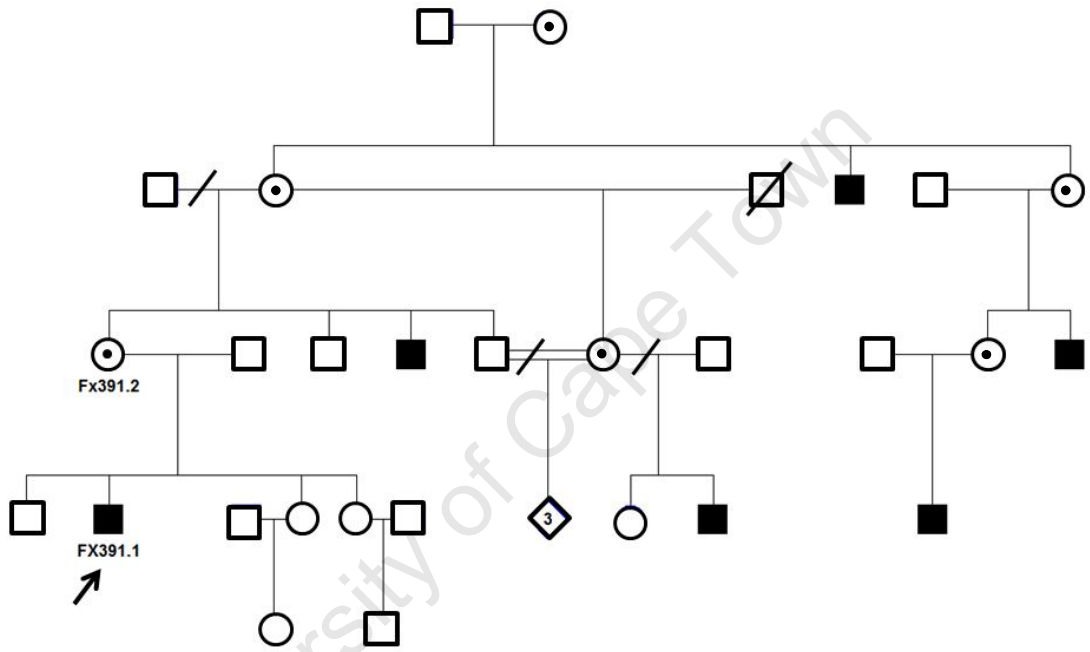


Figure 5.12 Pedigree of family Fx391

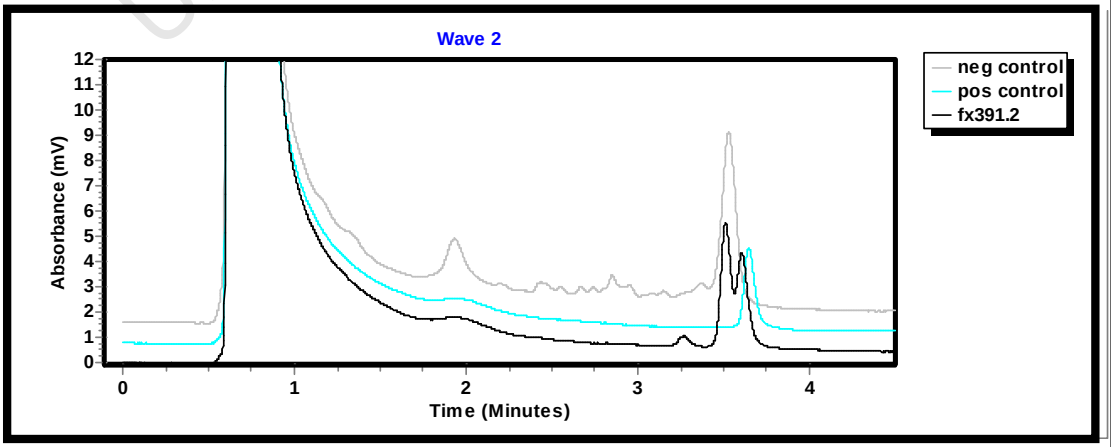


Figure 5.13 Chromatogram showing the c.428_451dup mutation in carrier (Fx391.2). Here PCR products of *ARX* exon 2a are separated on the basis of size using dHPLC. The c.428_451dup mutation positive (blue) elutes from the column later than the mutation negative (grey). The carrier mother (Fx391.2) (black) consisted of two peaks which correspond to the mutant and wild type allele.

Family Fx446 exhibited an X-linked inheritance of MR (Figure 5.14). The proband presented with mild-moderate NS-XLMR which included hypotonia and mild facial dysmorphism. The mother (Fx446.2) of the proband (Fx446.1) was the only individual available for carrier/mutation testing in the family. DHPLC analysis confirmed that the mother (Fx446.2) was a carrier of the c.428_451dup mutation (Figure 5.15).

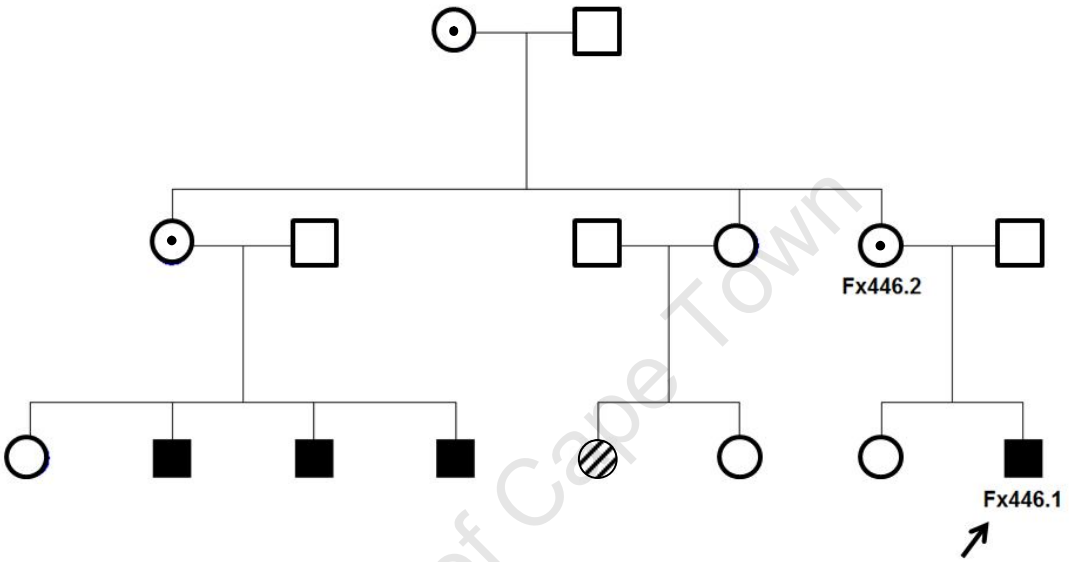


Figure 5.14 Pedigree of family Fx446. The proband has an affected female maternal cousin (hatched circle), however, this individual was not examined by clinicians participating in this study and therefore it was not known if the MR in this individual was due to the c.428_451dup.

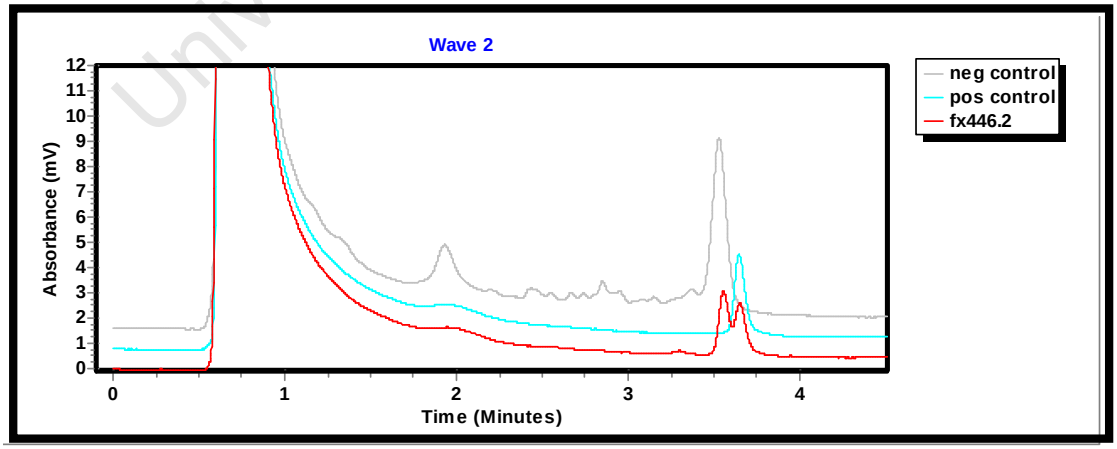


Figure 5.15 Chromatogram showing the c.428_451dup mutation in carrier (Fx446.2). Here PCR products of *ARX* exon 2a are separated on the basis of size using dHPLC. The c.428_451dup mutation positive (blue) elutes from the column later than the mutation negative (grey). The carrier mother (Fx446.2) has two peaks which correspond to the mutant and wild type allele (red).

The c.1119+79G>A variant in family Fx378

Fx378.1 was shown to carry an intron 3 variation designated, c.1119+79G>A (Figure 5.16). This intronic SNP has been previously reported (**rs2074002**) and has no known disease associations. While no population or individual data exists for this SNP it has been validated by multiple submissions to the NCBI database. The c.1119+79G>A was therefore not associated with XLMR in this family and was not investigated further.

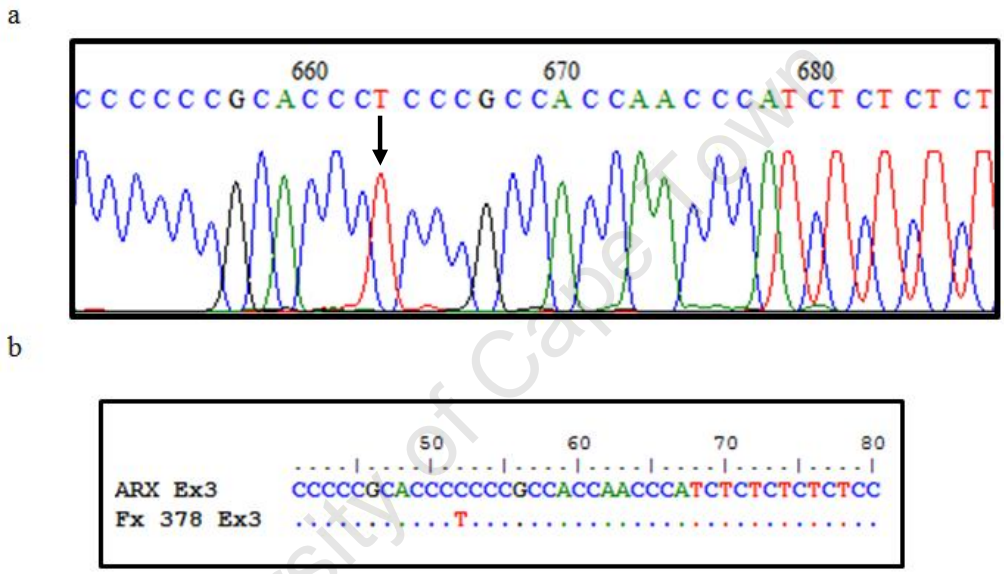


Figure 5.16 Chromatogram showing the c.428_451dup mutation in carrier (Fx446.2) Here PCR products of *ARX* exon 2a are separated on the basis of size using dHPLC. The c.428_451dup mutation positive (blue) elutes from the column later than the mutation negative (grey). The carrier mother (Fx446.2) has two peaks which correspond to the mutant and wild type allele (red).

5.3.1.3 The c.428_451dup and c.304ins(GCG)$_7$ PCR-based screen

Mutation screening for the c.428_451dup and c.304ins(GCG)$_7$ in 183 isolated male MR patients revealed no polyalanine expansions in this group of patients (Figure 5.17).

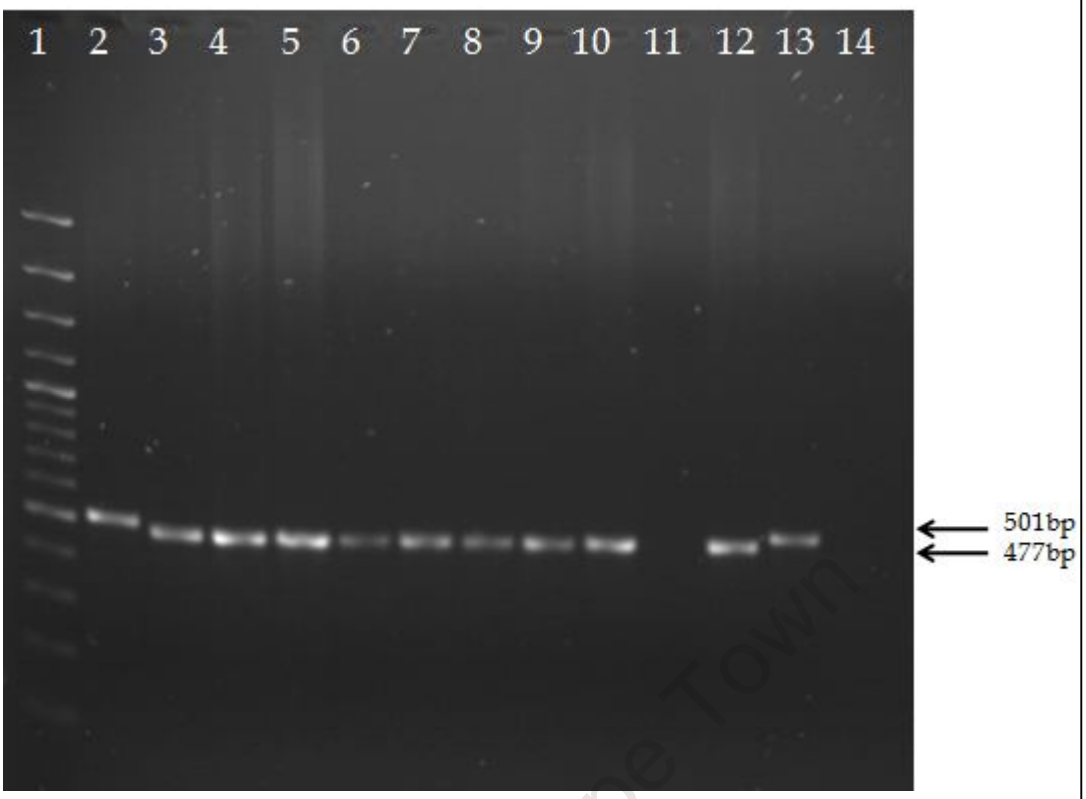


Figure 5.17 The *ARX* c.428_451dup and c.304ins(GCG)$_7$ PCR based screen in isolated males with MR. PCR products were assessed on a 3% agarose gel and visualised with EtBr. Lane 1 shows the 100bp molecular weight marker, Lanes 2 and 13 show samples previously demonstrated as positive for the c.428_451dup mutation (501bp), while lanes 3-12 illustrate a subset of the 183 isolated MR males in this screen (PCR of the sample in lane 11 failed), Lane 14 shows the no-DNA control. The c.428_451dup and c.304ins(GCG)$_7$ mutations was not detected in any of these individuals or the remainder of the cohort of 183 isolated MR males.

5.3.2 MUTATION DETECTION IN *KDM5C* USING DHPLC

5.3.2.1 dHPLC analysis in the 26 KDM5C exons

DHPLC analysis of the twenty *KDM5C* amplicons in the cohort of 25 patients identified three DNA sequence variations. Aberrant dHPLC profiles were identified in the samples corresponding to individuals Fx135, Fx361 and Fx277 in exons 4, 18 and 19 respectively (Figure 5.18).

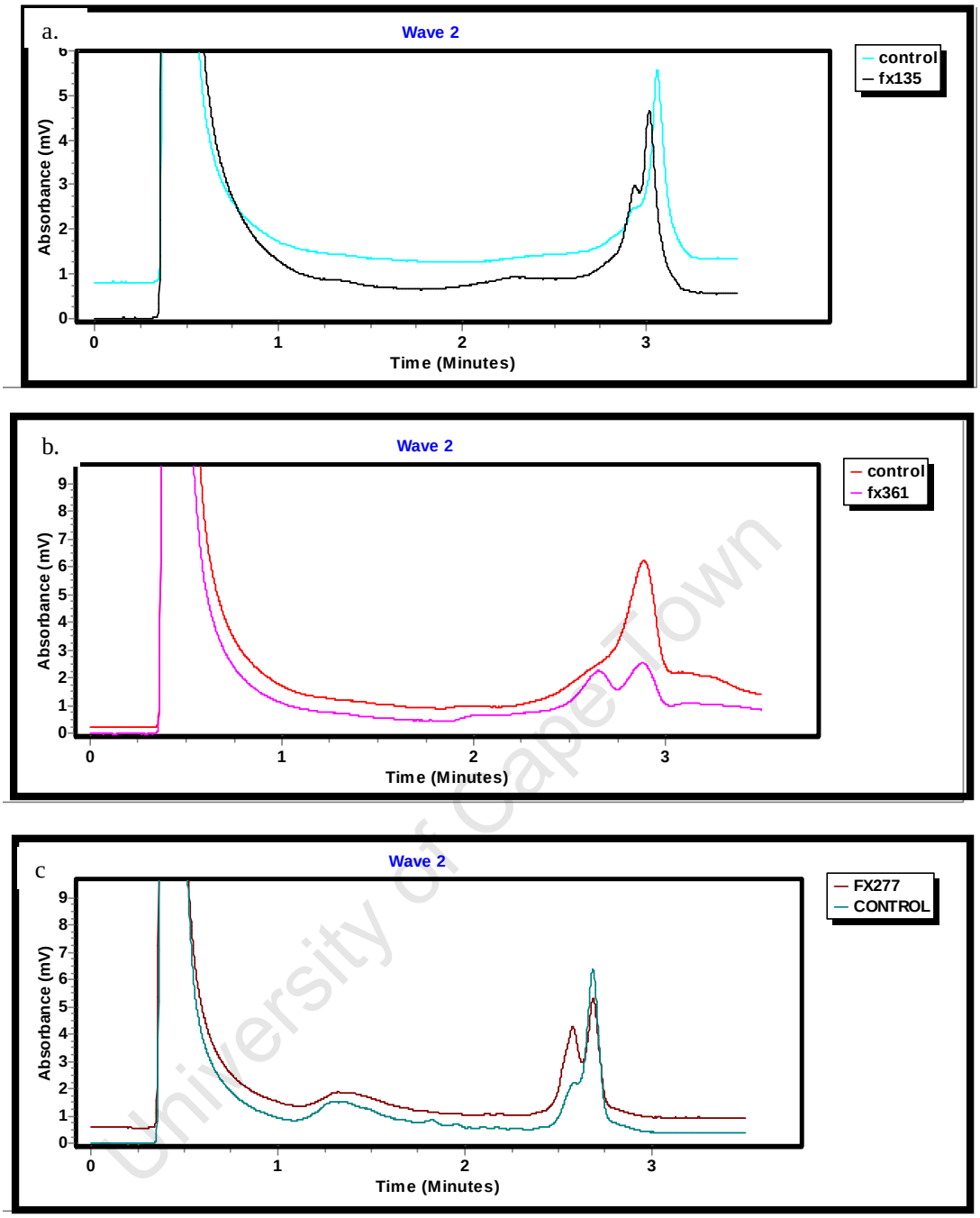


Figure 5.18 Chromatograms of the *KDM5C* variants. All PCR products were subject to dHPLC under semi-denaturing conditions. (a) The exon 4 variant in sample Fx135.1 . Here, Fx135.1 (black) elutes before the control (blue) and there is subtle evidence of two peaks indicative of heteroduplex formation. (b) The exon 18 variant in sample Fx361.1. Here, the presence of two peaks in sample Fx361.1 (pink) is evident as compared to the homoduplex in the control sample (red). DHPLC was performed at a temperature of 63.4°C (c) The exon 19 variant in Fx277.1. The heteroduplex formation indicative of an underlying sequence variation is evident by the presence of two peaks in Fx277.1 (brown) as compared to the homoduplex peak in the wildtype sample (grey) DHPLC was performed at a temperature of 63.4°C

5.3.2.2 Sequencing analysis of samples showing aberrant dHPLC profiles and variant screening in additional cohorts.

DNA sequencing of samples, Fx135, Fx361, Fx277, showing aberrant dHPLC profiles revealed the exact underlying nature of the sequence variation in amplicons of exon 4, 18 and 19 respectively.

The c.351-38C>T variant in family Fx135.

Individual Fx135 was shown to carry an intronic change (intron 3) designated c.351-38C>T (Figure 5.19). Due to the variant's intronic nature, this mutation has no direct affect on the amino acid sequence, however, a splicing alteration could not be ruled out at this point. This variation has not been previously described in any of the SNP databases. Unfortunately, no additional family members were available for segregation analysis, therefore the prevalence of the c.351-38C>T variant was assessed in a background population.

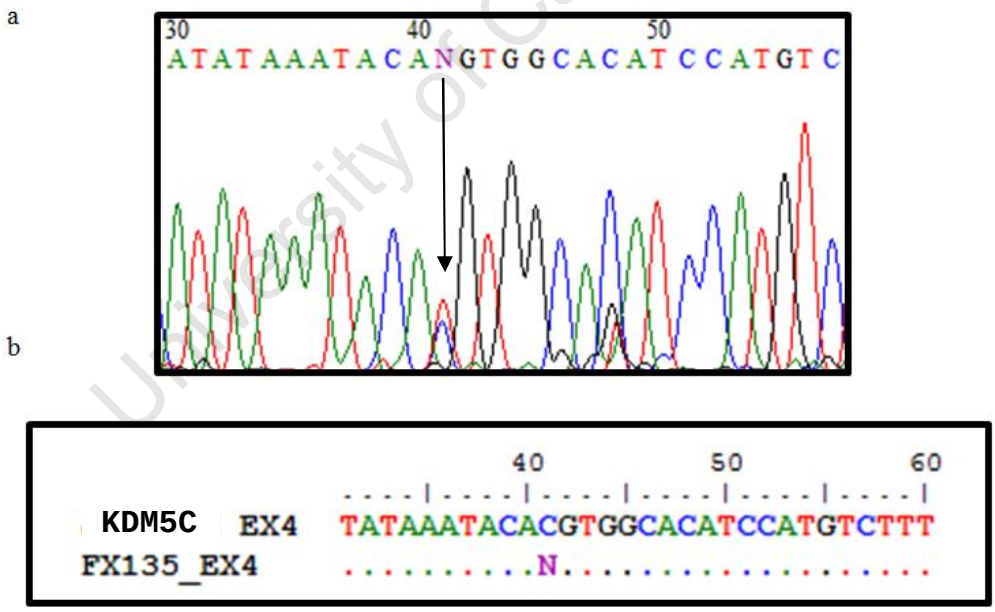


Figure 5.19 DNA sequence analysis results illustrating the c.351-38C>T in *KDM5C* intron 3 in individual Fx135.1. The sample Fx135.1 represents an affected female therefore the (a) electropherogram of Fx135.1illustrating the c.351-38C>T indicated by the black arrow and designated N owing to the presence of two alleles at this position. (b) An exert of the BioEdit alignment of the consensus sequence (top row) to sample Fx135.1 (bottom row) illustrating the conserved sequence (represented by the dots (.)) as well as the C>T mismatch (again represented as N).

The presence of c.351-38C>T variation was screened, using a BbrP1 digest, in 100 unaffected male controls that were matched for ethnicity. This variation was detected in three of the 100 DNA samples from unaffected males (Figure 5.20). Due to the presence of this intronic variation in unaffected individuals it was concluded that the c.351-38C>T was a novel SNP and therefore not associated with MR pathogenesis in family Fx135.

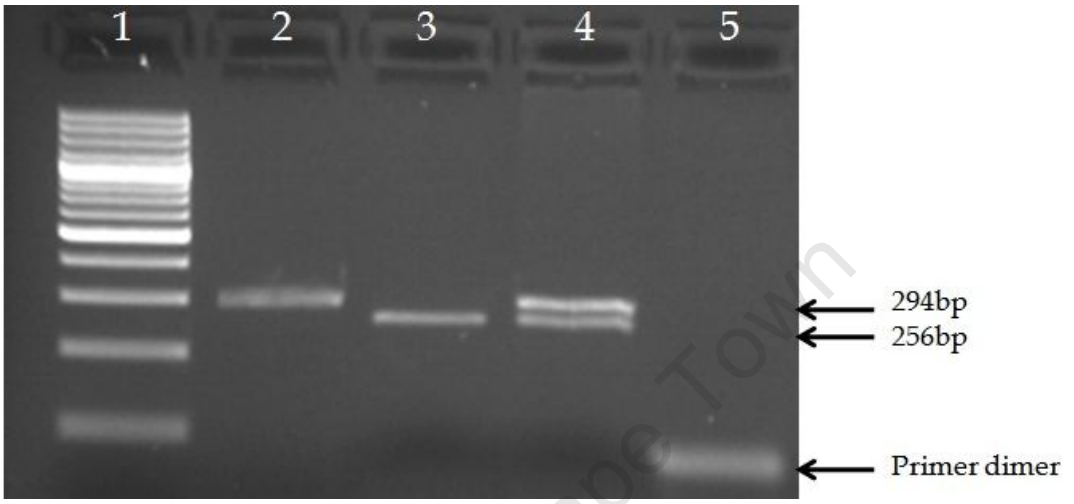


Figure 5.20 Screening for the *KDM5C* c.351-38C>T variant in 100 background controls. The presence of the c.351-38C>T variation destroys a BbrPI RE site at nucleotide 39 of the exon 4 PCR product. Therefore only in the presence of the wildtype sequence will cleavage occur (256bp), while the mutant sequence remains intact (277bp). Restriction enzyme digest products were separated on the basis of size using gel electrophoresis (2% agarose gel) and visualised with EtBr. Lane 1: molecular weight marker (100bp), Lane 2: male sample from background population positive for the c.351-38C>T variation, Lane 3: wildtype sample, Lane 4 sample Fx135.1 affected female heterozygous for c.351-38C>T, Lane 5: no-DNA control.

The c.2517-7_8insTAC variant identified in family Fx361

A c.2517-7_8insTAC in intron 17 was identified in individual Fx361 (Figure 5.21). This variant has been reported previously in a male patient with XLMR (Jensen et al. 2005). Jensen and colleagues (2005) did not detect the c.2517-7_8insTAC change in 312 control X-chromosomes screened. Furthermore, in this study, by ARMs-PCR it was demonstrated that this variant was not present in 100 ethnically matched unaffected males (Figure 5.22). Unfortunately, no additional family members were available for co-segregation analysis. However, given that the c.2517-7_8insTAC variation has now been identified in two XLMR individuals but not 412 control X chromosomes, the presence of this intronic variation was further assessed in the 76 probands of clear XLMR families by ARMS-PCR in an attempt to identify additional XLMR individuals with this intronic variant. Unfortunately, the c.2517-7_8insTAC was not detected in this cohort of MR patients.

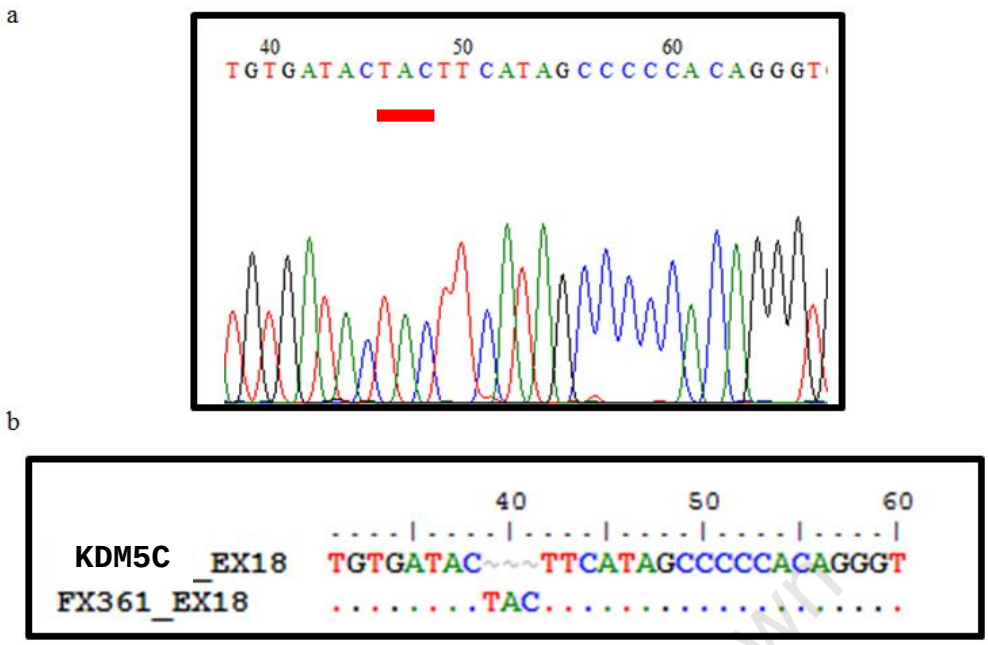


Figure 5.21 DNA sequence analysis results illustrating the c.2517-7_8insTAC in *KDM5C* intron 17 in individual Fx361.1. (a) An electropherogram of sample FX361.1, the c.2517-7_8insTAC variation is underlined in red. (b) An exert of the BioEdit alignment of the consensus sequence (top row) to sample Fx361.1 (bottom row) illustrating the conserved sequence (represented by the dots (.)) as well as the c.2517-7_8insTAC (underlined in red)

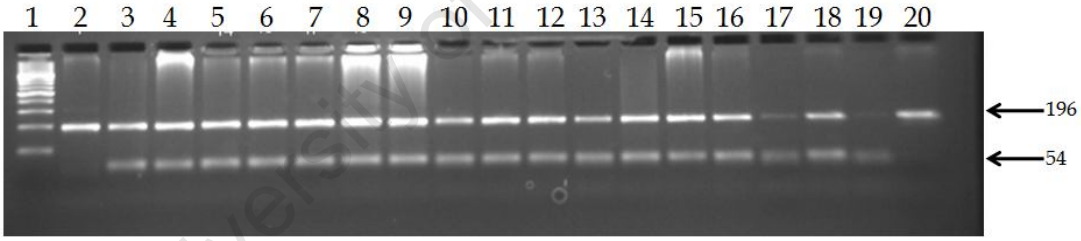


Figure 5.22 ARMS-PCR products to detect the c.2517-7_8insTAC variation in control X-chromosomes. PCR products were separated on the basis of size using gel electrophoresis (3%) and visualised with EtBr Lane 1: 100bp molecular weight marker, Lanes 2 and 20: sample Fx361.1 mutation positive for the c.2517-7_8insTAC variation and hence not producing the smaller 54bp band. Lanes 3-19: a subset of the 100 unaffected males screened, this variation was not detected in this background population group.

The c.2623-45G>A variant in family Fx277.

Sequencing in individual Fx277 revealed an intronic c.2623-45G>A change in intron 18 (Figure 5.23). This change has no direct consequence on the amino acid sequence but an effect on splicing could not be dismissed. This variation has not been previously described in either of the SNP databases (NCBI, Ensembl). Furthermore, no additional family members were available for co-segregation analysis, therefore screening for this variant in 100 ethnically matched unaffected males was attempted.

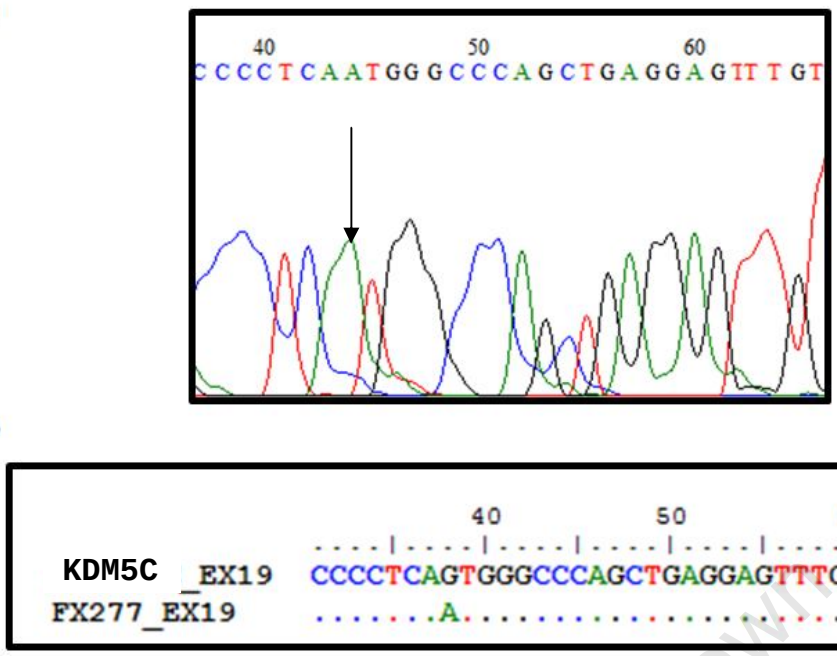


Figure 5.23 DNA sequence analysis results illustrating the c.2623-45G>A in *KDM5C* intron 18 in individual Fx277.1. (a) An electropherogram of FX277.1, the c.2623-45G>A transition is indicated by the black arrow. (b) An exert of the BioEdit alignment of the consensus sequence (top row) to sample Fx277.1 (bottom row) illustrating the conserved sequence (represented by the dots (.)) as well as the G>A mismatch.

The c.2623-45G>A could not be assessed in control X chromosomes using Dde1 restriction enzyme (RE) digest analysis. There are multiple Dde1 cleavage sites within this PCR fragment, with the wildtype having 6 sites compared to the five present in the c.2623-45G>A variant. Subsequent to Dde1digestion of the *KDM5C* exon 19 PCR product, it was predicted that the wildtype and mutant would produce bands of 16, 34, 66, 73, 93, 248 and 16, 48, 66, 73, 93, 248 respectively. The difference between the two banding patterns (the 34bp and 48bp products) could not be discerned by PAGE and therefore the c.2623-45G>A could not be screened in a control population using this method. Therefore, an ARMS-PCR was designed as an alternative to RE digestion. The ARMS-PCR was predicted to produce two PCR products (544bp and 57bp) in the wildtype, while only one band (544bp) in the presence of the c.2623-45G>A variation. However, despite several optimisation steps, the difference between the wildtype 57bp product and the primer dimer could not be discerned and the variant could not be assessed using this method. Due to the deeply intronic nature of this variation and thus the enhanced likelihood of the variation not playing a role in disease, this variant was not further assessed using a more expensive molecular technique such as dHPLC or DNA sequencing.

5.4 DISCUSSION

5.4.1 THE *ARX* GENE

Mutations in the *ARX* gene are thought to be one of the leading causes of XLMR accounting for up to 9.5% of this disorder (Poirier et al. 2006). The purported high prevalence of mutations makes *ARX* the second major contributor to XLMR after *FMR1*. For these reasons, *ARX* was an excellent candidate for mutation screening in affected individuals with X-linked inheritance who are negative for the *FMR1* 5'UTR expansion. In this study, mutation detection of the *ARX* gene coding region (and intron/exon boundaries) in 113 XLMR probands revealed two disease-causing c.428_451dup mutations in patients Fx391.1 and Fx446.1. In addition, a known SNP c.1119+79G>A (rs2074002) was identified in intron 3 of patient Fx378.1 and a novel SNP (c.300G>A) was detected in exon 2a of individual Fx362.1. Finally, no c.428_451dup or c.304ins$(GCG)_7$ expansions were detected in the 183 individuals screened by the allele fragment length-based PCR assay.

The c.428_451dup mutation detected in individuals Fx391.1 and Fx446.1 was originally described in 5 XLMR families and 1 sporadic MR case. The duplication segregated in these XLMR families and was not detected in 300 control X chromosomes, leading the authors to conclude that this duplication was responsible for the MR etiology in these families (Bienvenu et al. 2002). Since then the c.428_451dup has been reported in numerous cohorts and accounts for 45% of all mutations reported to date (Appendix 5A). In both families Fx446 and Fx391 described here, the c.428_451dup mutation segregated with the disorder and was therefore predicted to be the disease-causing mutation in these families.

A diverse range of phenotypes have been associated with this 24bp duplication including MRX, Partington syndrome and West syndrome (Bienvenu et al. 2002, Gronskov et al. 2004, Stromme et al. 2002a). However, the majority of c.428_451dup patients (32 of 38 cases (84%)) present with NS-XLMR. Concomitantly, patients from families Fx446 and Fx391 have a non-syndromic presentation of XLMR, with Fx391 presenting with mild facial dysmorphism and hypotonia.

The c.428_451dup mutation results in the expansion of the second ARX polyalanine tract from 12 to 20 alanine residues. The pathogenic nature of this mutation has not been clearly elucidated and its pathogenic role has been the source of conflicting results. Polyalanine expansions in other genes, particularly transcription factors (e.g. *ZIC2* and *HOXD13*) have been shown to exert their pathogenic effects by protein aggregation and increased apoptosis (Albrecht and Mundlos. 2005). Based on this premise it was hypothesised that the expansion of the second ARX polyalanine tract invoked pathogenic effects by a similar apoptosis-induced mechanism. Investigations to test this disease mechanism have reached contradictory conclusions. Two reports showed that c.428_451dup expansion does not cause nuclear protein aggregation (Poirier et al. 2004, Shoubridge et al. 2007), while unpublished results referred to in a review by Gecz and colleagues (2006) suggested that nuclear aggregates are a feature of this polyA expansion (Gecz, Cloosterman and Partington. 2006). Given the compelling experimental evidence that the c.428_451dup expansion does not result in aggregation of the mutant protein, compared to the unsubstantiated unpublished results, it is likely that nuclear aggregation is not the c.428_451dup disease-causing mechanism (Poirier et al. 2004, Shoubridge et al. 2007).

These data suggest that an alternative disease mechanism to protein aggregation is induced by the second polyalanine tract expansion. Polyalanine tracts have been hypothesised to be integral to the maintenance of protein tertiary structure which determines protein-protein interactions and/or DNA binding (Amiel et al. 2004). In support of this theory, polyalanine expansions in the transcription factor, ZIC2, affect the ability of this protein to bind target DNA sequences, resulting in virtually complete abrogation of target gene transcriptional activity (Brown et al. 2005). In a similar fashion, ARX polyalanine expansions may affect DNA-binding to target genes with concomitant effects on the proteins transcriptional repression and activation activity. Now that a number of ARX target genes have been identified by gene expression analysis in *Arx* knockout mice (Fulp et al. 2008), it will be possible to test this theory. The mRNA expression profile in c.428_451dup positive patients for the subset of identified ARX target genes could be assessed using a quantitative method (such as quantitative real-time PCR) in order to test for dysregulation at these genes.

Given the severe clinical presentation of patients with truncating and complete loss of function *ARX* mutations (i.e. XLAG and ACC) it is not expected that the c.428_451dup mutation, with its mild clinical manifestation, to lead to a complete loss of function. It would be interesting to compare target gene expression profiles between different *ARX* mutations so as to give an indication of the function of the various ARX conserved domains.

The known SNP (c.1119+79G>A) was idenitifed in intron 3 of individual FX378.1, was found to be catalogued in the reference SNP database curated by NCBI (accession number **rs2074002**). Therefore, this SNP cannot be associated with the disorder in family Fx378.1 due to its presence in the general (and unaffected) population.

A c.300G>A transition was detected in individual Fx362.1, this silent mutation has not been previously described in the SNP databases of either NCBI or Ensembl. Segregation analysis in the family demonstrated that this alteration did not track with the disorder in the family, as neither the proband's affected (Fx362.2) nor unaffected (Fx362.3) brother carry this variant. For these reasons, the c.300G>A alteration was not the disease-causing mutation in Family Fx362. Investigations in the background population would determine whether the c.300G>A is a rare variant or a novel SNP.

Interestingly, the c.300G>A variation was shown to disrupt the ESE site, SRp55, using the ESEFinder prediction tool described in Chapter 4 (Cartegni et al. 2003). The disruption of this site could lead to aberrant intron splicing which while not the disease-causing change in this family could contribute to the disease pathogenesis. Similarly Gronskov and colleagues (2004) detected an *ARX* polymorphism c.1347C>T in MR patients as well as unaffected family members. This polymorphism created a novel, strong Srp55 binding site, leading the authors to speculate that this change could act as a modifier (Gronskov et al. 2004). Given the low mutation detection rates in XLMR families, particularly in sib-pairs, it has been hypothesised that a proportion of familial XLMR recurrence can be accounted for using a polygenic model (Mandel and Chelly. 2004, Raymond and Tarpey. 2006, Ropers. 2006). These X-linked modifiers could collectively contribute to a genetic

predisposition toward learning disability. It is hypothesised that these two putative splicing variants (c.300G>A and c.1347C>T) could act as modifiers, but mRNA analysis should be conducted first to test for aberrant splicing.

Furthermore, in support of the 'X-linked risk allele' theory, an additional variant was detected in the same sib-pair of family Fx362 in a collaborative study between the Wellcome Trust Sanger Institute in Hinxton, Cambidge and the Division of Human Genetics UCT. In this study, individual Fx362.2 was shown to carry a truncating variant (p.S76fsX7) in *ARSF* (Tarpey et al. 2009). This variant did not segregate in the family, with the affected brother (Fx362.1) testing negative for this truncating mutation. In addition, *ARSF* truncating variants were identified in five of 1346 controls, leading to the conclusion that ARSF was not an XLMR gene. Furthermore, truncating variants in additional X-linked genes were detected in normal controls, with these deleterious mutations being present in up to 1% of X-linked genes in the general population (Tarpey et al. 2009). The accumulation of truncating variants in X-linked genes (X-linked risk factors) above a certain threshold is an attractive inheritance model for the 'unexplained' XLMR families, which requires further investigation.

While the role of X-linked 'risk-factors' in the pathogenesis of MR remains to be elucidated, the role of the *ARX* polyalanine expansions is quite clear. Cumulatively, expansions of the first two polyalanine tracts account for 61% (Appendix 5A) of all *ARX* mutations. For these reasons the presence of these polyalanine expansions was assessed in 183 males with isolated, unexplained MR. While no mutations were detected in this group, this result was in keeping with previous reports suggesting a low mutation frequency in certain cohorts, particularly those including sporadic MR cases. Gronskov et al. (2004) screened 682 males with developmental delay for the two common polyalanine expansions and identified only two disease-causing mutations (~0.3%) (Gronskov et al. 2004). Also, the EuroMRX consortium reports only a 0.1% detection rate in sporadic MR (de Brouwer et al. 2007).

Mutation detection in the *ARX* gene posed numerous problems, primarily due to the excessive GC content of the *ARX* coding regions. PCR amplification of exon 2a which harbours a 78% GC content was the most troublesome. The PCR was

eventually optimised using a specialised buffer and a high fidelity Taq polymerase. The difficulties experienced, however, prompted an investigation of the relationship between GC content and PCR efficiency in collaboration with Prof H. Viljoen and colleagues at the Department of Chemical and Biomedical Engineering, University of Nebraska-Lincoln, USA. Through these studies it was established that the PCR efficiency of GC-rich templates is strongly dependant on PCR annealing time, with a shorter annealing time (3-6 seconds) being optimal (Mamedov et al. 2008).

In this South African study, two c.428_451dup mutations were identified from a total of 113 clear or putative XLMR families assessed, resulting in a mutation detection frequency of ~1.8% in this cohort. A figure much lower than previous reports of *ARX* mutations accounting for between 6.6-9.5% of XLMR cases (de Brouwer et al. 2007, Mandel and Chelly. 2004, Poirier et al. 2006). However, of the 113 XLMR patients assessed here, 36 were part of a sib-pair (29 brother pairs, 7 brother-sister pairs), 34 showed putative X-linked inheritance and only 43 were clearly XLMR (Appendix 5C). Therefore, *ARX* mutations account for 4.6% of clear XLMR cases in this South African cohort, a figure much closer to previous reports. Also, the detection rate of 0/29 brother pairs (0%) was similar to that of the EuroMRX consortium of 1.5% (de Brouwer et al. 2007).

The two *ARX* c.428_451dup mutations detected in this South African cohort reflected current *ARX* literature, in terms of both mutation detection rates and clinical presentation. It was therefore concluded that diagnostic testing for the two polyalanine expansions using the PCR and gel electrophoresis-based approach described here was both necessary and justified in the South African setting. Patients with an X-linked history of NS-XLMR or a known *ARX* MR 'epilepsy' syndrome should be investigated for these polyalanine expansions subsequent to 5'UTR *FMR1* expansion testing.

In keeping with these conclusions, diagnostic testing for the first two *ARX* polyalanine expansions (c.428_451dup and c.304ins$(GCG)_7$) has been implemented at the Division of Human Genetics, UCT. Since the test's inception 12 patients have been referred to the Division by local clinicians and to date, one c.304ins$(GCG)_7$ insertion has been identified. Incorporating this test into the diagnostic protocol has

improved the diagnostic yield of patients with MR and will continue to do so. This enhanced diagnostic yield in turn affords improved genetic management for MR families, positively impacting the burden of disease alleviation,particularly important in a developing country such as South Africa.

5.4.2 THE *KDM5C* GENE

The *KDM5C* gene is one of the more commonly mutated XLMR genes, with reported prevalence of 4.2% and 4.3% in XLMR and sib-pair cohorts respectively (de Brouwer et al. 2007). In general, patients with *KDM5C* mutations present with NS-XLMR, although common additional clinical features speech impairment (45%), short stature (40%), hyperreflexia or spasticity (35%), aggressive behaviour (30%), epilepsy/seizures (30%) and hand abnormalities (30%) (Abidi et al. 2008, Adegbola et al. 2008, Jensen et al. 2005, Rujirabanjerd et al. 2009, Santos et al. 2006, Tzschach et al. 2006) In this study, 25 XLMR patients with one or more of these features were screened for mutations in the 26 *KDM5C* exons but no disease-causing mutations were identified.

While no disease-causing *KDM5C* mutations were identified, three patients were shown to carry intronic variants. The first, in patient Fx135.1, designated c.351-38C>T and residing in intron 3, was not previously described in the SNP databases. This novel change was detected in 3/100 (3%) of the ethnically matched male control DNA samples analysed by allele-specific restriction enzyme digest. It was therefore concluded that the c.351-38C>T variant constitutes a novel SNP.

The second alteration (c.2517-7_8insTAC) in intron 17 of *KDM5C* was identified in patient Fx361.1. The proband presents with MRX and speech delay as does his younger brother but there was no further family history. This insertion (c.2517-7_8insTAC) has been previously described in an XLMR patient, although the pathogenic nature is unknown (Jensen et al. 2005). In an attempt to elucidate the association between this intronic insertion and MR the presence of the alteration was assessed in 100 ethnically matched male controls by ARMS-PCR but did not detect the change. Interestingly, Jensen and colleagues (2005) did not detect this variant in 312 control X chromosomes in their study either (Jensen et al. 2005).

The c.2517-7_8insTAC is located seven base pairs from the KDM5C intron 17/exon 18 boundary. Also, this insertion is located just 2bp upstream of the branch site. This branch site is rich in pyrimidines and is the binding site for the U2AF^{65} component of the splicesome complex (Cartegni, Chew and Krainer. 2002) (Figure 5.24). A disruption to the DNA tertiary structure by the c.2517-7_8insTAC insertion could compromise U2AF^{65} binding with concomitant effects on accurate splicing of *KDM5C* exon 18. To test this hypothesis it was first necessary to first confirm the c.2517-7_8insTAC insertion in the proband's affected brother, and secondly to sequence the patients mRNA to detect putative aberrant splicing. Unfortunately, despite multiple attempts, the family could not be re-contacted and thus it was not possible to perform these analyses.

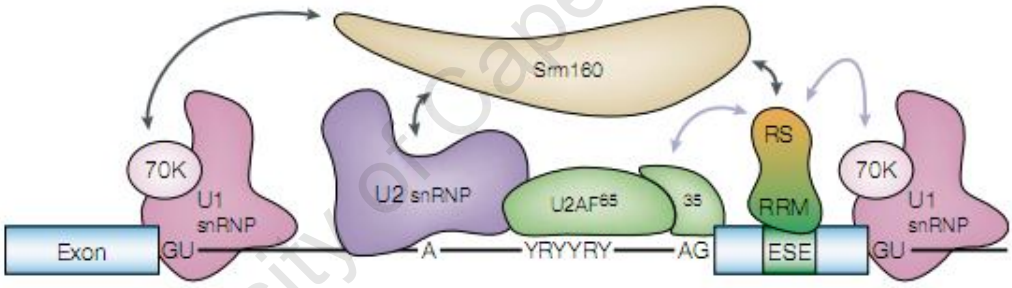


Figure 5.24 Assembly of the splicesome component, U2AF^{65}, at the pyrimidine-rich branch site. The consensus DNA sequence comprising the branch site is YRYYRY, for *JARID1C* intron 17 the consensus sequence reads TGTGAT. However, the flanking sequence in the wildtype reads TGTGATact (flanking sequence in lowercase) while the c.2517-7_8insTAC mutation alters the sequence to TGTGATtacact, this modification in DNA structure may affect the binding of U2AF^{65} with subsequent aberrant splicing of intron 17. (Figure from Cartegni, Chew and Krainer. 2002)

By ARMS-PCR, 77 probands from XLMR families were screened in an attempt to identify additional XLMR patients with this insertion, however, the c.2517-7_8insTAC was not detected. In conclusion, the detection of the c.2517-7_8insTAC in two affected individuals but not 412 (cumulative number) control samples, together with the insertion's close proximity to critical splicesome recognition sequences suggest that this alteration plays a role in MR pathogenesis but requires further validation.

The third intronic variant identified in this *KDM5C* mutation analysis was the c.2623-45G>A transition detected in patient Fx277.1. This intron 18 variant has not been described in any of the SNP databases. The presence of this variant in an ethnically matched male control population was not possible, as optimisation of the molecular techniques (allele-specific restriction enzyme digestion and ARMS-PCR) were unsuccessful. In addition, re-establishment of contact with this XLMR family was not possible, thus co-segregation analysis of the c.2623-45G>A transition could not be performed. While other methods (eg dHPLC and DNA sequencing) are available to detect this variant in a background population, they are prohibitively expensive. Given these limitations the exact role of this variant in the pathogenesis of MR remains unknown, however, given the variant's deeply intronic nature it was deemed unlikely to cause alterations in splicing of the *KDM5C* transcript.

To date, two large scale *KDM5C* mutation detection analyses have been reported in XLMR cohorts. The first was conducted by the European EuroMRX consortium which reported mutations in seven (4.2%) of the 166 XLMR families and five (4.3%) of the 116 brother-pairs screened (de Brouwer et al. 2007). The second report, from an American study, identified *KDM5C* mutations in two (8.6%) out of 23 linked XLMR families, one (0.6%) of the 172 XLMR families and one (1.1%) of 92 sporadic MR patients presenting with short stature (Abidi et al. 2008). By comparison, the study described in this dissertation included 25 MR patients, hence even though this cohort was enriched based on clinical presentation it was not unexpected that no mutations were detected. However, success in similar small cohorts has been reported elsewhere, *KDM5C* mutation detection in a cohort of 24 XLMR patients revealed a single disease-causing mutation (Santos et al. 2006). Therefore, despite no mutations being detected in this cohort, screening of *KDM5C* in a larger cohort of patients with the characteristic clinical presentation would give a better indication as to the prevalence of mutations in this XLMR gene in South African MR patients.

5.4.3 SUMMARY AND CONCLUSIONS

The higher mutation detection rates in certain XLMR genes make them excellent candidates for molecular screening investigations in clinically stratified cohorts. The 450 probands in this study were originally referred to the Division of Human Genetics, UCT for molecular testing of the most common XLMR mutation, the *FMR1* 5'UTR expansion. Of these referrals ~23% of XLMR families are positive for this Fragile X syndrome mutation, a prevalence in keeping with international reports.

To identify the XLMR etiology in the remaining 77% of families *ARX*, the next most frequently mutated XLMR gene was screened, as well as *KDM5C*, commonly implicated in the disorder. While the mutation detection frequencies in these studies were low, 4.6% and 0% respectively, it should be noted that the cohorts investigated were relatively small in comparison to other studies reporting higher frequencies. Therefore, extended mutation screening in these XLMR genes would give a more accurate reflection of these genes' mutation frequencies in the South African population.

Given the relatively low detection rates and the vast number of XLMR genes, it remains difficult to stratify cohorts for screening in XLMR genes despite emerging genotype-phenotype correlations aiding clinical stratification. This is especially true in a developing country such as South Africa with fewer resources and a battling public health system. In the future it is envisaged that the cost of microarray and high-throughput DNA sequencing will reduce and the screening of all XLMR genes will be relatively simple and cost effective. However, this reality is a long-way off in South Africa and in the meantime, additional diagnostic strategies need to be put in place.

Despite the low pick-up rates in these XLMR genes, studies such as these are crucial in establishing molecular techniques which will extend the repertoire of diagnostic genetic testing offered to XLMR patients and their families. To this end, the diagnostic testing service for the two most common *ARX* mutations (c.428_451dup

and c.304ins$(GCG)_7$) has been successfully implemented under the auspices of the NHLS.

The central hypothesis of this dissertation was to ascertain the effect of 'epigenetic' XLMR gene mutations on the DNA methylation pattern of affected individuals. In this chapter, two patients with mutations in the *ARX* gene have been identified, a transcriptional regulator, these affected individuals are thus potential candidates for whole-genome DNA methylation analysis.

PART A

CONCLUSIONS

IDENTIFICATION OF DISEASE-CAUSING MUTATIONS IN SOUTH AFRICAN XLMR PATIENTS

The central aim of this dissertation is to assess the epigenomic profiles of patients with known disease-causing XLMR gene mutations. In order to achieve this aim it was necessary to first identify patients with these XLMR gene mutations within the 450 MR families that comprise the Division of Human Genetics, UCT, DNA bank. The complex genetic landscape of MR dictated that several molecular approaches be employed in order to achieve this objective. These included: CNV analysis (Chapter 2), linkage analysis (Chapter 3), positional candidate gene screening (Chapter 4), and finally functional candidate gene screening (Chapter 5).

CNV detection using whole-genome SNP arrays led to the detection of two disease-causing mutations. The first, a 17p11.2Del in individual Fx343.1, causes the Smith Magenis syndrome in this MR patient. This heterozygous deletion spans approximately 3Mb and encompasses over 30 genes. While many of the genes in this region may potentially play a role in epigenetic regulation it would be difficult to delineate the potential effect of each gene in the event that there was a change in the epigenomic profile. Second to this, the heterozygous deletion detected here was autosomal, while the focus of this study was primarily concerned with X-chromosomal genes. Therefore, the 17p11.2Del was not a good candidate for investigation in the epigenetic phase of this dissertation.

The second disease-causing mutation was the de novo 0.8Mb Xq25Dup detected in an isolated case (XMR8.1) of MR. This duplication encompassed four known genes (*GRIA3*, *THOC2*, *XIAP*, *STAG2*) of which *GRIA3* has previously been implicated in XLMR. The *GRIA3* protein product is a glutamate receptor and is unlikely to affect

epigenetic regulation. The remaining genes have not been previously implicated in disease; however, *STAG2* represents an interesting candidate for contribution to the MR phenotype in this patient given the proteins putative functioning as a chromatin remodeler. Second to this, duplications in other chromatin remodelers (*ATRX* and *MECP2*) have been shown to cause MR. However, given the uncertain role of *STAG2* in the pathogenesis of the disorder in this individual, as well as the potential confounding effects of the additional duplicated genes, the Xq25Dup was not selected for follow-up epigenomic investigations.

By employing the linkage analysis technique in four XLMR families, the disease-associated X-chromosomal region was refined sufficiently for the selection of positional candidate genes in two of these families (Fx67 and XMR2). Subsequent positional candidate gene screening in family Fx67 entailed the mutation screening of the XLMR genes *ATP6AP2* and *TSPAN7*. However, no mutations were identified in either of these two XLMR genes. In addition, given the genetic heterogeneity of XLMR it was not possible to effectively prioritise mutation screening of the remaining 25 genes located within the disease-associated region in this family. Therefore, the disease-causing mutation in family Fx67 remains unknown at this stage. Conversely, linkage analysis and subsequent positional candidate gene analysis in family XMR2 led to the detection of a disease-causing mutation (c.5987_6011del) in the XLMR gene, *ATRX*. Given the proposed role of ATRX as a chromatin remodeler, patients with mutations in this gene were good candidates for the proposed epigenomic studies and investigation of this dissertation's central aim (Part B).

Notwithstanding this finding, further investigations were pursued in order to identify additional XLMR gene mutations for inclusion in the second phase of this study. Therefore, the functional candidate genes *ARX* and *KDM5C* were selected for mutation screening in clinically stratified cohorts on the basis of previously reported mutation frequencies. Screening of the second most commonly-mutated XLMR gene, *ARX*, led to the identification of the common *ARX* disease-causing mutation, c.428_451dup, in individuals Fx446.1 and Fx391.1. Subsequently, mutation screening for the two most common *ARX* mutations (c.428_451dup and c.304ins$(GCG)_7$) was incorporated into the diagnostic protocols offered for MR

patients by the NHLS based at the Division of Human Genetics, UCT. Since implementation of this diagnostic test, an additional patient with the c.304ins$(GCG)_7$ mutation has been identified. ARX has been proposed to function as a transcriptional regulator and therefore could potentially affect the epigenomic profile in mutation positive individuals (Part B). Mutation screening in *KDM5C*, also a significant contributor to XLMR, revealed no disease-causing mutations in the study cohort. While *KDM5C* mutation positive patients remain good candidates for epigenomic profiling given the protein's role in histone demethylation, these investigations were not possible in this study as no *KDM5C* mutations were detected.

In summary, the strategic molecular approaches adopted in this study have been successfully implemented in the identification of disease-causing mutations in the *ARX* and *ATRX* genes. Given the purported function of these genes as a transcription activator/repressor (*ARX*) and chromatin remodeler (*ATRX*), these genes both posed good candidates for DNA methylation analysis in mutation positive patients (Part B).

PART B

INVESTIGATION OF DNA METHYLATION PROFILES IN PATIENTS POSITIVE FOR MUTATIONS IN XLMR GENES

A large proportion (22%) of XLMR genes are either known or postulated to play a role in transcriptional regulation (Figure 1.3). These 'epigenetic' XLMR genes include transcription activator/repressors (e.g. *ARX*) and chromatin remodelers (e.g. *ATRX*). The functional roles of these protein families overlap in a co-ordinated way to control gene expression. Furthermore, the interaction of these protein families with the two primary epigenetic marks, histone modifications and DNA methylation, induce a chromatin state which is applicable to the transcriptional state required. Thus, it was hypothesised in this study that mutations in XLMR genes involved in this transcription regulation pathway would lead to aberrant DNA methylation.

In order to test this hypothesis, first, mutations in the *ARX* and *ATRX* genes were identified using various molecular approaches in a cohort of South African XLMR patients (Part A). Subsequently, these genes were prioritised for whole-genome DNA methylation analyses in mutation positive patients.

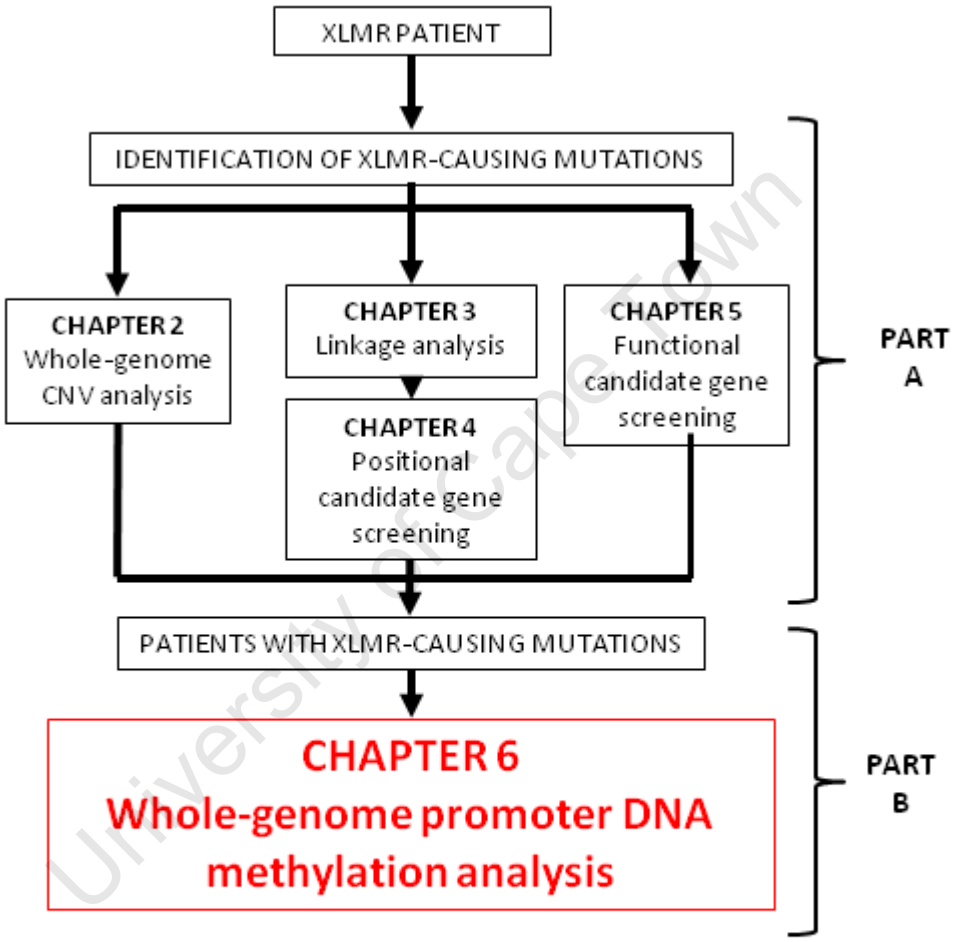
The *ARX* gene has been demonstrated to function as a transcription repressor and activator (Chapter 5), therefore this gene was a good candidate for whole-genome DNA methylation analysis. Despite, the identification of three mutation positive *ARX* patients, as well as the suitability of *ARX* to test this study's hypothesis, these patients were not further assessed for DNA methylation aberrations. The reasons for *ARX* exclusion were two-fold. Firstly, the expression profile of the *ARX* transcript suggest that the gene is not expressed (Stromme et al. 2002b), or else it is expressed at low levels in peripheral blood leucocytes (PBLs) (Ohira et al. 2002). Given that these cells were the only source of DNA available for this study, at this stage it was a concern whether impairment of ARX function would exert its pathogenic effects in this cell type. Secondly, large amounts of DNA were required for the whole-genome

DNA methylation experiments described in Chapter 6, these quantities of DNA were not available and difficulties were experienced in contacting the affected families.

The *ATRX* gene has been shown to function as a chromatin remodeler (Chapter 4). Also, unlike *ARX*, the *ATRX* transcript was shown to be expressed in adult PBLs (Villard et al. 1997). In addition, ATR-X patients exhibit α-thalassaemia due to the reduced expression of the α-globin gene (Gibbons et al. 1995), suggesting that ATRX functions in multiple tissues. Therefore, while DNA isolated from brain tissue would be ideal, it was hypothesised that lymphocytic DNA could be used to investigate DNA methylation profiles in these patients given the multitude of tissues affected in ATR-X syndrome. Finally, sufficient DNA from four affected males and three unaffected members of family XMR2 was available for whole-genome DNA methylation analysis. The *ATRX* mutation positive family, XMR2, was therefore used to test the hypothesis of this study.

CHAPTER 6

WHOLE GENOME DNA METHYLATION ANALYSIS



6.1 INTRODUCTION

The pathogenic mechanisms by which mutations in XLMR genes lead to cognitive deficits in MR patients are poorly understood. Thus, an understanding of which biological pathways are disrupted by mutations, as well as the intricate details of the resulting pathophysiology is essential if the ultimate goal of MR therapeutic intervention is ever to be realised. With this in mind, in this study, the role of DNA methylation as a pathogenic mechanism by which XLMR gene mutations exert their effects was investigated.

The *ATRX* gene is one of the 22% of XLMR genes which play a role transcriptional regulation. The participation of ATRX in gene regulation is supported by several lines of evidence, including:

- Association with chromatin remodelling complexes by interaction with proteins such as:
    - The repressive polycomb group protein, EZH2 (Cardoso et al. 1998). EZH2 is a polycomb group (PcG) protein involved in homeotic transcriptional repression during development
    - The heterochromatin protein, HP1 (McDowell et al. 1999)
    - Transcription/apoptosis cofactor, DAXX (Ishov, Vladimirova and Maul. 2004)
    - The MBD protein, MeCP2 (Nan et al. 2007)
- DNA hypomethylation at the acrocentric chromosomal regions which encompass the rDNA gene arrays in *ATRX* mutation positive patients. Conversely, DNA hypermethylation at Y-specific repeats (DYZ2) is also evident in these patients (Gibbons et al. 2000)
- The protein's localisation to PML-NBs, sites of transcription regulation (Xue et al. 2003)
- The weak chromatin remodelling ability of ATRX, as demonstrated by mononucleosome and triple-helix DNA displacement activity (Xue et al. 2003)

- Dysregulation of numerous genes involved in neurogenesis and neuronal development in conditional *Atrx* knock-out mice (Levy et al. 2008)
- The localisation of ATRX to an $H3K9_{me2}$-hotspot upstream of the Xist locus which controls X-inactivation in females (Baumann and De La Fuente. 2009).

Collectively, this evidence strongly motivates for a chromatin remodelling function for ATRX. Also, the aforementioned observations have led to the hypothesis that ATRX forms part of a neuronal transcription repression complex which targets methylated promoter regions (Figure 6.1) (Kramer and van Bokhoven. 2009).

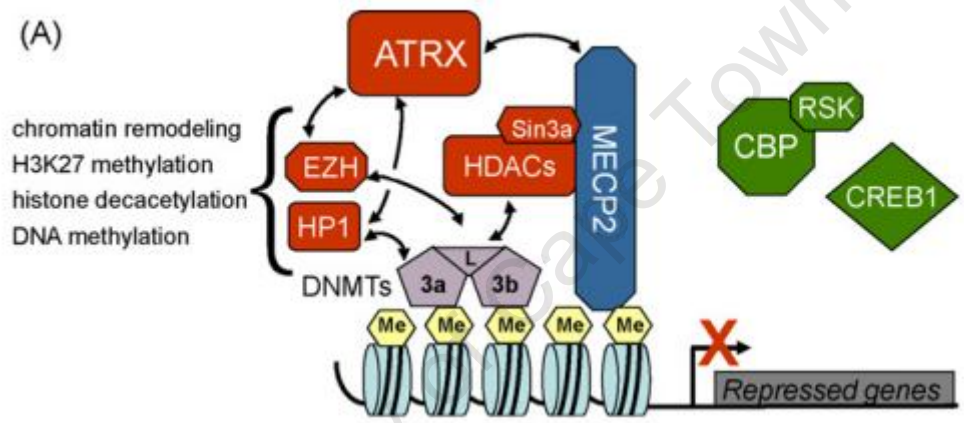


Figure 6.1 A theoretical transcription repression complex functioning in neuronal cells. The formation of this complex involves the targeting of repressive chromatin remodelers (ATRX, Sin3a, HDACs, EZH, HP1) to methylated DNA by interaction with MeCP2. This repressive complex could subsequently prevent the formation of the CBP-RSK-CREB1 transcription activation complex, thus inducing gene repression (Kramer and van Bokhoven. 2009).

While aberrations in DNA methylation have already been demonstrated in ATR-X patients at rDNA gene regions and Y-specific repeats (Gibbons et al. 2000), no whole-genome investigations of DNA methylation status in ATR-X patients have been conducted to date. Furthermore, given the substantial evidence for the role of ATRX in gene regulation and the concomitant importance of DNA methylation in this process, it was hypothesised that aberrant DNA methylation was a pathogenic mechanism associated with *ATRX* mutations. Therefore, in this study a whole-genome CpG island and promoter DNA methylation analysis in *ATRX* mutation positive family, XMR2, was conducted so as to assess the extent of DNA methylation aberration in these patients and potentially identify *ATRX* target genes.

6.2 METHODS

6.2.1 METHYLATION-DEPENDENT IMMUNOPRECIPITATION (MEDIP)-CHIP ASSAY

Methylation dependent immunoprecipitation and array hybridisation (MeDIP-Chip) was employed to assess the whole-genome promoter DNA methylation profiles of *ATRX* mutation positive patients, as compared to unaffected family members (procedure outlined in Figure 6.2).

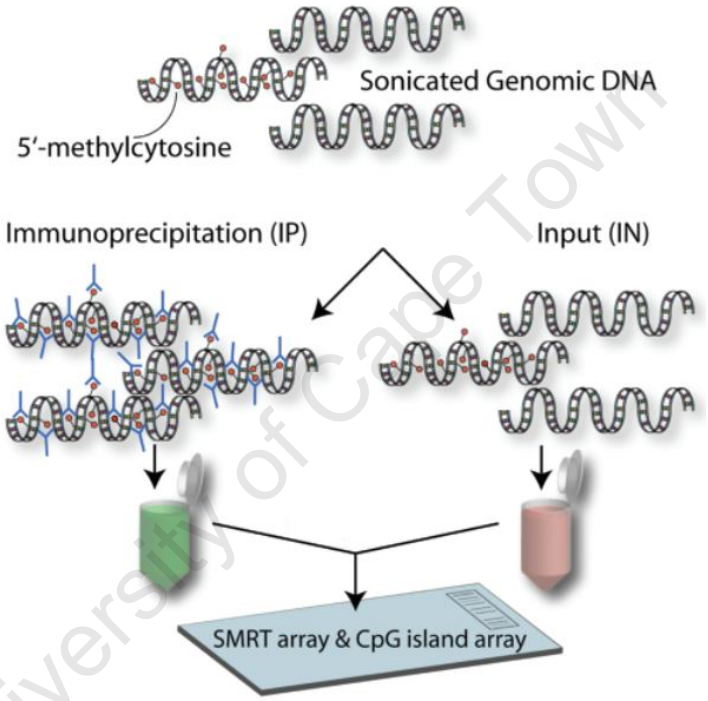


Figure 6.2 The whole-genome MeDIP-Chip procedure adopted in this study. DNA samples are sonicated to shear the DNA into smaller fragments. Immunoprecipitation using a 5-methylcytosine specific antibody enriches the DNA for methylated portion of the genome (IP). Subsequently, the IP and input (IN) DNA are differentially fluorescently labelled and, in this study, hybridised to a CpG island and promoter array in order to generate DNA methylation profiles for each individual. (Figure from Wilson et al. 2006).

Given the lack of resources and expertise available at UCT, these investigations were conducted together with Dr. Andrew Sharp at the Department of Human Genetics and Development, University of Geneva Medical School, Geneva, Switzerland.

6.2.1.1. MeDIP

DNA samples obtained from the four *ATRX* mutation positive patients (XMR2.5/13/25/26) as well as three unaffected male family members (XMR2.14/17/21) (pedigree in Appendix 3A) were subject to a MeDIP protocol. MeDIP was used in order to enrich DNA samples for the methylated CpG portions of the genome. This was achieved by immunocapture of the methylated fraction of the genome using an antibody specific for 5-methyl-Cytosine (Weber et al. 2005).

This MeDIP protocol entailed the sonication of 15µg of each DNA sample to obtain fragment sizes of 200-800bp using a Branson 450D sonifier. Subsequently, 2.5µg of sonicated DNA was removed from the solution and set aside for array labelling and hybridisation (untreated DNA fraction). The remainder of the sonicated DNA was denatured and subject to immunocapture by incubation with 10µg of 5meth-C antibody [Diagenode] at 4°C with rotation overnight. Protein A sepharose beads [Life technologies] were used to bind the DNA-antibody component of the solution. After a brief incubation (2 hours at 4°C with rotation), the DNA-antibody-agarose complex was washed thrice with IP buffer in order to retain only the DNA-protein complex. The protein A agarose beads-antibody complex was subsequently digested using proteinase K [Qiagen] (incubation at 55°C with rotation overnight). A phenol-chloroform DNA extraction was performed in order to purify the methylated CpG fraction of DNA, which was subsequently precipitated using 100% ethanol, 5M NaCl and glycogen, followed by washing with 70% ethanol. The precipitated DNA was hereafter resuspended in TE buffer and quantified on a Nanodrop [ThermoScientific].

6.2.1.2 Array preparation, labelling and hybridisation

In order to obtain a genome-wide promoter DNA methylation pattern, DNA (untreated and methylation-enriched fractions) from all patients and controls were analysed by array hybridisation using the HD2 deluxe CpG island and promoter array [NimbleGen]. This commercially-available array platform consists of 2.1 million 50-75mer probes which encompass all annotated promoters, CpG islands and miRNA promoters, as well as manually selected ENCODE regions. These regions are covered at a 100bp resolution or less [NimbleGen].

The array hybridisation procedure entailed DNA labelling by random priming using Cy3 (methylated DNA fraction) and Cy5 (sonicated, untreated DNA)-conjugated random nonamers, as per manufacturer's instructions [Nimblegen]. Equal quantities (35µg) of Cy3 and Cy5 labelled DNA were combined and added to the reagents of the NimbleGen Hybridisation kit (according to manufacturer's specifications). Samples were then hybridised to the HD2 deluxe CpG island and promoter array [NimbleGen] by incubation in a MAUI hybridisation station [BioMicro systems] for two days. Subsequent to DNA sample-array hybridisation, the arrays were washed using the Nimblegen Wash Buffer kit according to manufacturer's instructions.

All arrays were scanned on the G2565 Agilent scanner [Agilent Tecnologies] at a 5µm resolution according to manufacturer's recommendations. The NimbleScan Version 2.5 [NimbleGen] software was used to analyse raw data and generate independent probe log2 fold changes. The SignalMap V.1.9 [NimbleGen] software was used to visualise log2 fold changes according to genomic location of the probes. Duplicate array-hybridisations and analyses were performed for each DNA sample.

6.2.1.3 Statistical analysis for differentially methylated region (DMR) detection

In order to identify potential DMRs, statistical analyses were performed to establish which loci harboured differential methylation ratios between the ATR-X patients (XMR2.5/13/25/26) and controls (unaffected family members) (XMR2.14/17/21). All statistical analyses were performed by Dr. Eugenia Migliavacca (Department of Human Genetics and Development, University of Geneva Medical School, Geneva, Switzerland) using software form the Bioconductor project (Gentleman et al. 2004).

Prior to statistical analyses a quantile normalisation was performed for each array, this normalisation accounts for systematic bias incurred when comparing different arrays (e.g. DNA-labelling, MeDIP enrichment) (Bolstad et al. 2003). Also, a novel approach to outlier replacement was implemented (A. Sharp, unpublished data) and a linear smoothing function was also applied (Pelizzola et al. 2008).

Subsequent to these normalisation and filtering steps, a moderated t-score test was performed (as employed in the LIMMA software package Smyth. 2004) in order to detect significantly different methylation values between the ATR-X mutation positive (XMR2.5/13/25/26) and control individual (XMR2.14/17/21) groups. In order to account for multiple testing, a false discovery rate (FDR) correction was applied (Benjamini and Hochberg. 1995). Potential DMRs were identified by clusters of probes with significantly differential methylation fold changes at a relatively low stringency threshold. These clusters were defined as regions which encompassed a single probe with a FDR adjusted p-value ≤ 0.01 that were flanked by at least two additional probes with nominal p-values ≤ 0.01. These clusters were subject to further computational analyses and filtering methods.

6.2.1.4 Computational analysis of potential DMRs

All overlapping potential DMRs were merged in order to form a non-redundant set of genomic loci using the online software tool, Galaxy (http://main.g2.bx.psu.edu/). In addition, chromosomal regions harbouring CNVs and segmental duplications (SDs) are prone to alterations in copy number which can produce false positives in array data (A. Sharp, unpublished data, Vega et al. 2009). Thus, Galaxy was used to remove all potential DMRs that overlapped with known CNVs (Conrad et al. 2009) and segmental duplications (dataset from the UCSC genome browser, http://genome.ucsc.edu/). Finally, subsequent to these filtering techniques, Galaxy was used to retrieve the genomic DNA sequence of putative DMRs.

6.2.1.5 Prioritisation of potential DMRs for validation

In order to validate the results of the MeDIP-Chip experiments, a number of DMRs were selected for bisulphite DNA sequencing, these included putative DMRs with:

- the highest absolute value t-score statistic (four loci)
- the highest average (for all differentially methylated probes in that region) absolute t-score statistic (four loci)

6.2.2. VALIDATION OF PUTATIVE DMRs BY BISULPHITE-DNA SEQUENCING

In order to validate the potential DMRs identified by this MeDIP-Chip assay, Sanger DNA sequencing was utilised, subsequent to the treatment of genomic DNA with sodium bisulphite. Treatment of genomic DNA with sodium bisulphite results in the conversion of unmethylated cytosine residues to uracil by deamination, while methylated cytosines are protected from this conversion by the 5-methyl-C modification. Target DMRs were then amplified using the bisulphite-treated DNA as a template and amplicons were subject to DNA sequencing. Subsequently, unmethylated cytosine residues appear as thymine nucleotides on the resulting electropherogram, which were clearly distinguishable from the cytosine residues (which correspond to methylated Cs).

6.2.2.1 Bisulphite treatment of DNA

In order to convert unmethylated cytosine residues to uracil nucleotides, patient and control DNA was treated with the Epitect® bisulfite kit [Qiagen] according to manufacturer's instructions.

6.2.2.2. PCR of putative DMRs

PCR primers were designed to amplify each of the potential DMRs using the Methyl Primer Express software V1.0 [Applied Biosystems] (Appendix 6A). PCR amplification was performed in a 25µl reaction containing 100ng of bisulphite-treated DNA, 10µM of each forward and reverse primer, 2.5µM dNTP's [Bioline], 1× JumpStart REDtaq buffer [Sigma-Aldrich] and 1U JumpStart REDtaq DNA polymerase [Sigma-Aldrich]. A 'touchdown' PCR profile was employed, entailing an initial denaturation step at 95ºC for 5 min, followed by 10 cycles of 95ºC for 10s, 66ºC for 10s (with 1ºC decrements/cycle) and 72ºC for 30s. Subsequently, 30 cycles of 95ºC for 10s, 55ºC for 10s and 72ºC for 30s were performed and a final elongation at 72ºC for 5 min completed the PCR amplification. PCR was conducted on a Px2 thermal cycler [Thermo Electron Corporation] and PCR success was assessed by electrophoresis on a 1.5% agarose gel and visualised with EtBr.

6.2.2.3 DNA sequencing of putative DMRs

Subsequent to PCR amplification, each putative DMR was subject to DNA sequencing. Each PCR product was incubated with 5U Exonuclease 1 [Fermentas] and 1U Shrimp Alkaline Phosphotase [Fermentas] (ExoSAP treatment) per 10µl of PCR product to remove unincorporated dNTPS and primers from the PCR reaction. The PCR product and enzymes were incubated at 37ºC for 30min and was followed by enzyme inactivation at 80ºC for 15min. The purified PCR products were then subject to DNA sequencing as described before (Section 4.2.1.5).

6.2.3 QUANTIFICATION OF DNA METHYLATION AT DMRS

Those loci which were found to harbour true positive DMRs by bisulphite DNA sequencing (Section 6.2.2.3), were cloned and individual colonies sequenced in order to obtain a semi-quantitative measure of DNA methylation across the region in both the ATR-X patients and unaffected controls.

6.2.3.1. Cloning of DMR PCR product

PCR was performed for each of the validated DMRs as above (Section 6.2.2.2) in the four affected males (XMR2.5/13/25/26) and the three unaffected family members (XMR2.14/17/21). PCR products were purified using the Qiagen Gel Extraction Kit as per manufacturer's instructions [Qiagen]. Purified PCR products were cloned by T/A cloning using the pGEM®-T Easy Vector System [Promega]. This cloning protocol involved the insertion of the purified DMR PCR product into the pGEM®-T Easy Vector [Promega] at the synthesised T/A cloning region within the multiple cloning site (Figure 6.3) according to manufacturer's instructions [Promega]. Subsequently, the vector-insert recombinant DNA molecules were transformed into chemically competent JM109 E.coli cells using a heat-shock protocol [Promega]. Cells were incubated in 950µl 2× YT broth at 37°C with shaking (~150rpm) for 90 min. Thereafter, 200µl of cells were spread onto plates containing 2× YT agar supplemented with 100µg/ml ampicillin [Sigma-Aldrich], 100µg/ml of bromo-chloro-indolyl-galactopyranoside (X-gal) [Roche] and 200µg/ml Isopropyl β-D-1-thiogalactopyranoside (IPTG) [Roche]. Plates were incubated overnight at 37°C.

The pGEM-T Easy vector carries the *lacZ* gene, which encodes the α fragment of β-gal, while JM109 cells contain the β fragment of β-gal, association of the two fragments results in the formation of a functional β-gal enzyme. This active β-gal enzyme hydrolyses X-gal to produce a blue dye resulting in blue colonies, in the presence of IPTG. Thus, upon ligation of the insert into pGEM-T Easy, the α fragment of β-gal is disrupted and thus β-gal is not produced, resulting in the formation of white colonies. Therefore, blue colonies contain only vector, whilst white colonies contain vector plus the DMR PCR product. Hence, white colonies containing the insert were selected for DNA sequencing.

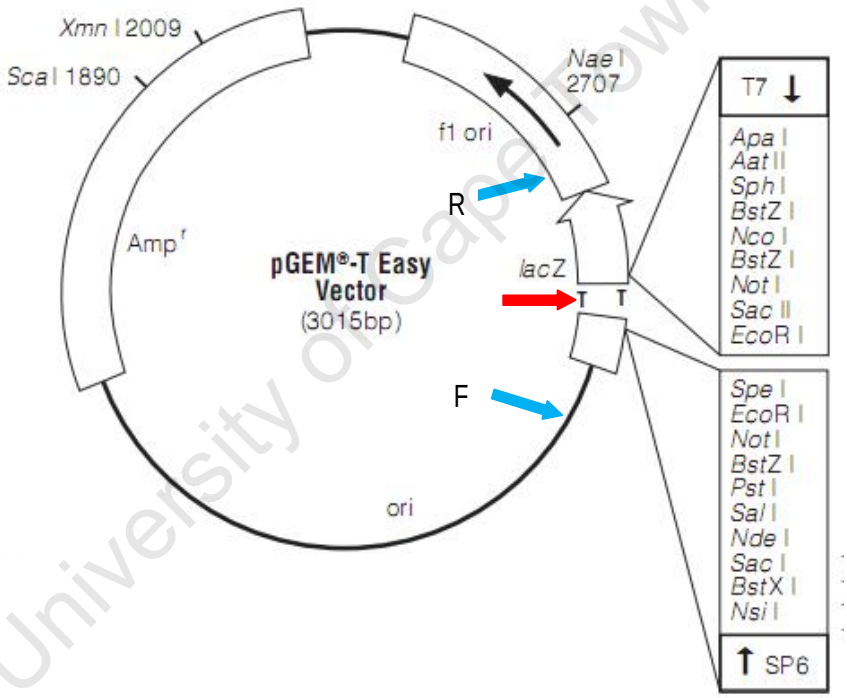


Figure 6.3 The pGEM®-T Easy Vector [Promega] circle map. DMR bisulphite-treated DNA PCR products were inserted at the T/A cloning site (red arrow). PCR was performed using the pUC/M13 forward (F) and reverse primers (R) (blue arrows). DNA sequencing was performed using the pUC/M13 forward primer.

6.2.3.2 PCR and DNA sequencing of cloned DMR products

Twenty positive (white) colonies generated by T/A cloning in each patient and control individual of family XMR2 were selected for DNA sequencing. These colonies were picked and placed in 10µl distilled water, incubated at 95°C for five minutes and centrifuged briefly at 13000rpm for 45s. Thereafter, 2µl of this solution was used as the template for the PCR reaction. In addition to this plasmid DNA, the PCR reaction was performed using 10µM of the plasmid-specific [pGEM®-T Easy Vector] forward (5' gttttcccagtcacgac 3') and reverse primers (5'caggaaacagctatgac 3'), 2.5µM dNTP's [Bioline], 1.5µM $MgCl_2$ [Applied Biosystems], 1× AmpliTaq Gold® PCR buffer [Applied Biosystems] and AmpliTaq Gold® DNA polymerase [Applied Biosystems] in a final reaction volume of 25µl. The PCR cycling conditions and gel electrophoresis were performed as above (Section 6.2.2.2.). DNA sequencing was performed after ExoSAP treatment (Section 6.2.2.3) as described previously (Section 4.2.1.5) using the aforementioned plasmid-specific forward primer (Figure 6.3). The software programme BiQ Analyzer was used to analyse the DNA sequencing and generate 'lollipop' diagrams (Bock et al. 2005).

6.3 RESULTS

6.3.1 WHOLE-GENOME CPG ISLAND AND PROMOTER DNA METHYLATION ARRAY ANALYSIS

Comparison of whole-genome promoter DNA methylation profiles between four affected and three unaffected individuals of family XMR2, led to the identification of 112 unique putative DMRs (Appendix 6B). These putative DMRs were selected on the basis of one probe exhibiting a FDR-adjusted p-value ≤0.01, combined with at least two additional flanking probes with a nominal p-value ≤0.01. Subsequent to the removal of all DMRs containing overlapping regions with known CNVs and SDs, a total of 107 unique putative DMRs were identified.

From these 107 putative DMRs, a total of eight loci were selected for validation. These candidate DMRs (Table 6.1) were selected as they possessed the highest absolute t-score value (rows 1-5, Table 6.1) or the highest absolute value average t-score (rows 1,4,6-8, Table 6.1) i.e. where the average t-score was calculated using t-

score values for all probes exhibiting significantly differentially methylation values. Interestingly, the chromosomal region encompassing probe, CHR16FS010387819 (row 4, Table 6.1 and Figure 6.4/5), scored highest at both a single probe, as well as for the average across the DMR.

Table 6.1 The eight DMRs selected for validation by bisulphite DNA sequencing

	Probe ID	t-score value	P-Value	FDR* adjusted P-Value	AVG t-score value	AVG FDR adj p-value
1.	CHR16FS010387819	9.1126	2.49E-18	2.66E-12	5.367544	5.08E-02
2.	CHR11FS126375528	-8.1971	2.47E-15	4.77E-10	-4.34842	2.38E-01
3.	CHR17FS034606720	-8.1907	2.59E-15	4.77E-10	-4.43846	1.51E-01
4.	CHR05FS001853381	7.8763	2.45E-14	1.87E-09	4.940175	6.36E-02
5.	CHR02FS065138632	6.8146	2.98E-11	5.03E-07	4.176412	9.56E-02
6.	CHR19FS015436362	-6.8001	3.26E-11	5.43E-07	-5.30631	5.53E-03
7.	CHR18FS031177845	-7.0706	5.77E-12	1.58E-07	-4.98466	3.93E-02
8.	CHR05FS074359693	5.8915	7.40E-09	2.91E-05	4.8544	3.32E-03

Highlighted in red, the criterion upon which selection for validation was based

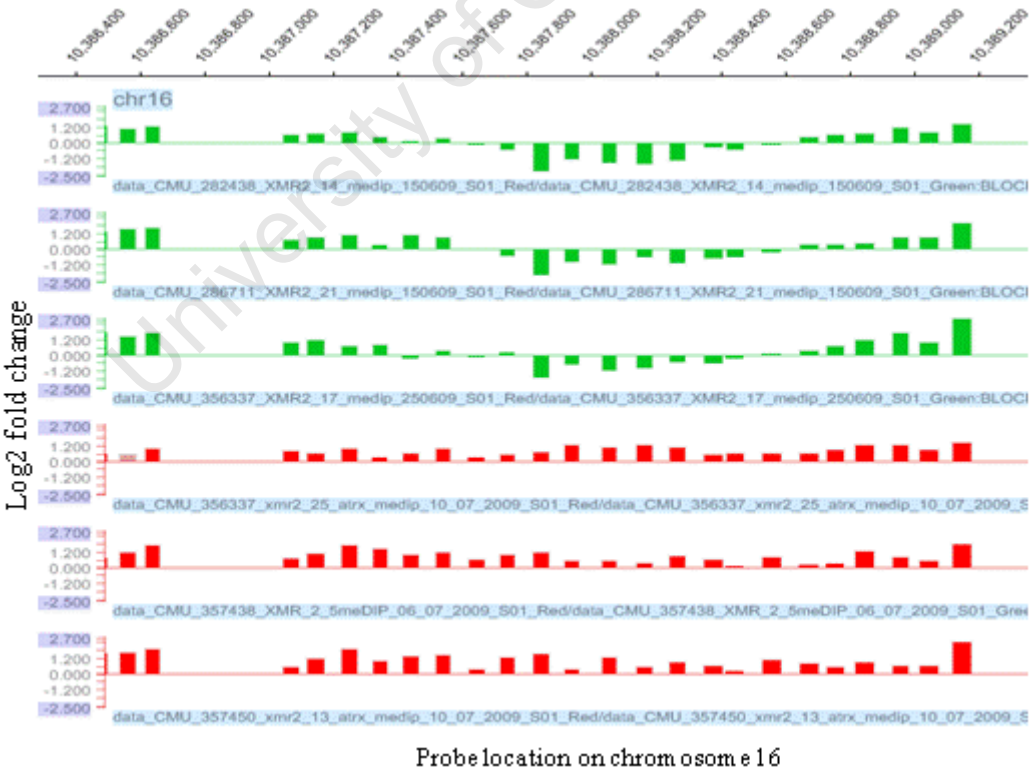


Figure 6.4 The DMR encompassing the probe CHR16FS010387819; a DMR identified by whole-genome promoter DNA methylation analysis in *ATRX* mutation positive patients. The signalMap [Nimblegen] output illustrating the hypermethylation at the chromosome 16 (~10.387Mb – 10,388Mb) region in *ATRX* mutation positive patients (red) as compared to the unaffected sibs (green) of family XMR2.

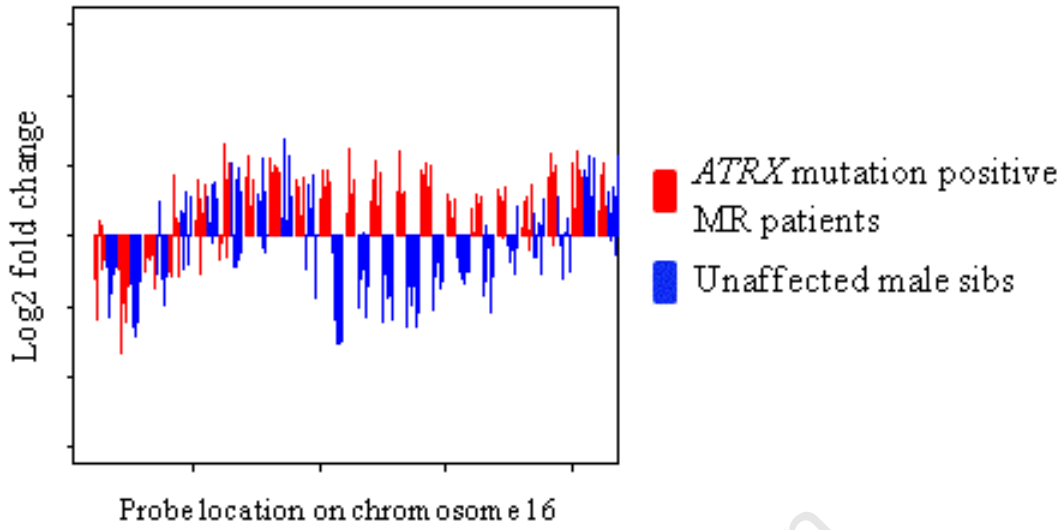


Figure 6.5 The DMR encompassing the probe CHR16FS010387819; a DMR identified by whole-genome promoter DNA methylation analysis in ATRX mutation positive patients and subsequent statistical analyses. The output generated by statistical analysis also showing the hypermethylation of this 16p region in the ATR-X patients (red) as compared to the unaffected sibs (blue).

6.3.2 VALIDATION OF PUTATIVE DMRS

Bisulphite DNA sequence analysis of the eight putative DMRs prioritised for validation revealed one true positive DMR, encompassing the central probe CHR16FS010387819 (Figure 6.6a, b). The DMR corresponds to an approximately 1kb interval on chromosome 16 (16p13.2) (Figure 6.6c). This chromosomal interval encompasses a CpG island which corresponds to the promoter region of the *Activating Transcription Factor 7 Interacting Protein 2 (ATF7IP2*). While the electropherogram generated from these DNA sequencing investigations validated the hypermethylation at this locus, this technique did not allow the quantification of the proportion of cells with methylation. Therefore, a semi-quantitative DMR PCR cloning and DNA sequencing approach was implemented across the *ATF7IP2* promoter region.

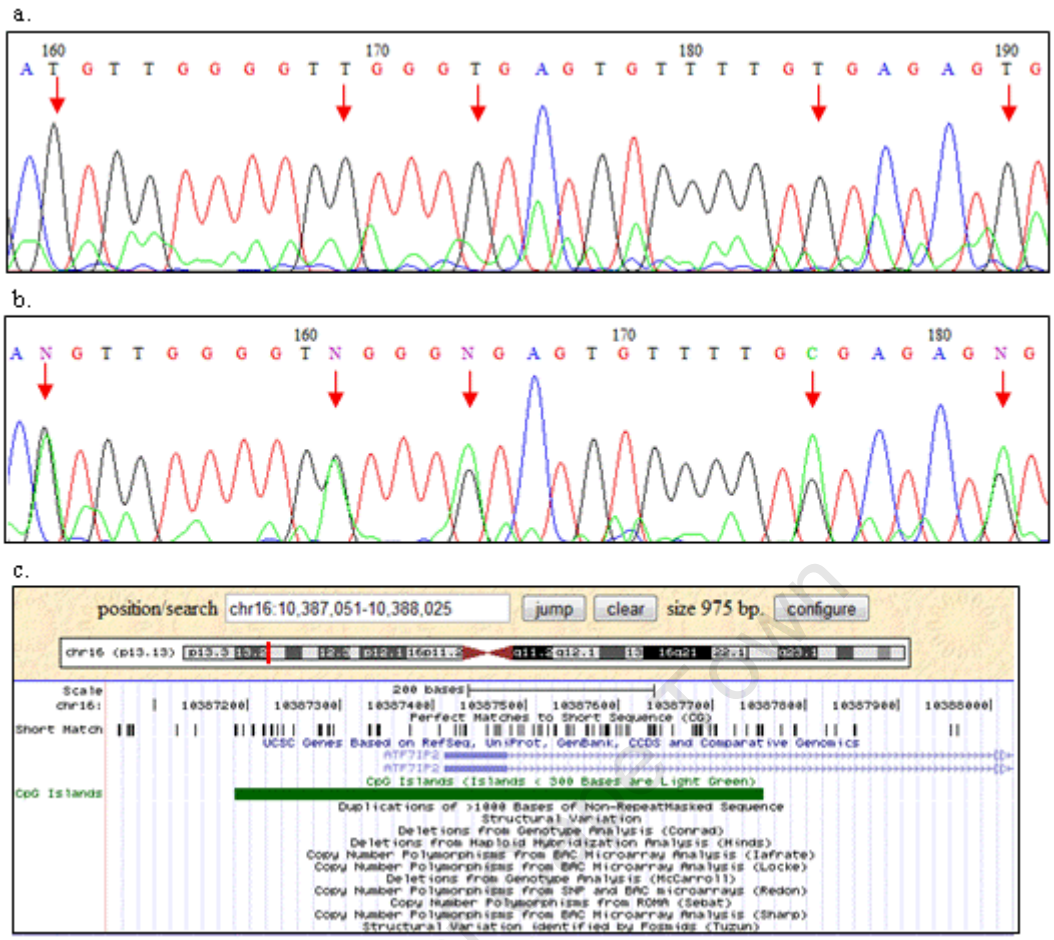


Figure 6.6 The 16p13.2 DMR verified by bisulphite DNA sequencing in affected individuals of family XMR2. (a) An electropherogram of the bisulphite sequencing results of the *ATF7IP2* promoter region in an unaffected individual XMR2.14. The red arrows indicate the location of the CpG sites, which in this individual correspond to thymine (T) residues, illustrating the complete conversion of unmethylated cytosine residues by the bisulphite treatment. (b) An electropherogram of the same 16p13.2 locus in affected individual (XMR2.25), here the red arrows indicate the presence of unmethylated cytosine residues (T) as well as the methylated cytosine residues (C) at the CpG sites. This electropherogram indicates *ATF7IP2* hypermethylation in this patient and confirms the results of the DNA methylation array results. (c) A UCSC exert of the hypermethylated *ATF7IP2* locus in the ATR-X patients, this hypermethylation occurs in the CpG island upstream of *ATF7IP2*. No segmental duplications or CNVs are present in this region.

The remaining seven putative DMRs prioritised for validation by bisulphite DNA sequencing showed no discernable difference between patients and controls at the CpG sites investigated (Figure 6.7). Therefore, it was concluded that these seven putative DMRs represent a false positive result.

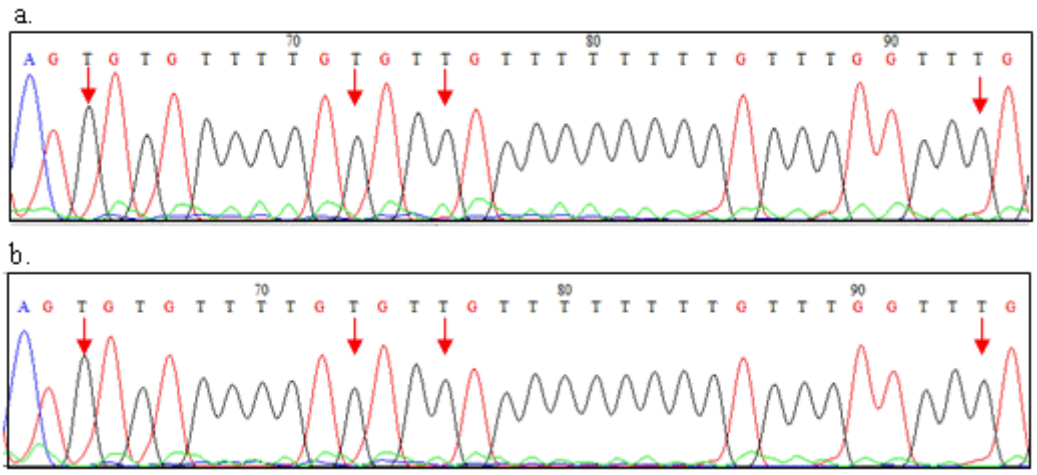


Figure 6.7 The DMR encompassing the CHR17FS034606720 probe, an example of a DMR identified by the DNA methylation whole-genome promoter array that was not confirmed by bisulphite DNA sequencing. (a) The electropherogram generated by bisulphite DNA sequencing in the unaffected individual (XMR2.17). Red arrows correspond to the CpG sites; all sites contain only thymine residues, illustrating a complete lack of cytosine methylation at this methylation sensitive locus. (b) The corresponding electropherogram from affected individual (XMR2.25) illustrates the same unmethylated cytosine residues (T nucleotides) at all CpG sites. Thus there is no differential DNA methylation at this particular locus in affected individuals of ATR-X family XMR2.

6.2.3 QUANTIFICATION OF DNA METHYLATION AT THE *ATF7IP2* PROMOTER REGION

The semi-quantitative DMR PCR cloning investigations showed variable methylation ratios across 23 CpG sites of between 8-34% in the ATR-X patients (XMR2.5/13/25/26) and 0-7% in the unaffected family members (XMR2.14/17/21) (Figure 6.8). Also, the distribution of methylated cytosine residues demonstrated intercellular and inter-individual variability (Figure 6.9/10/11). These results show that *ATRX* mutation positive individuals have significantly increased levels of DNA methylation at the *ATF7IP2* promoter region as compared to the unaffected members of family XMR2.

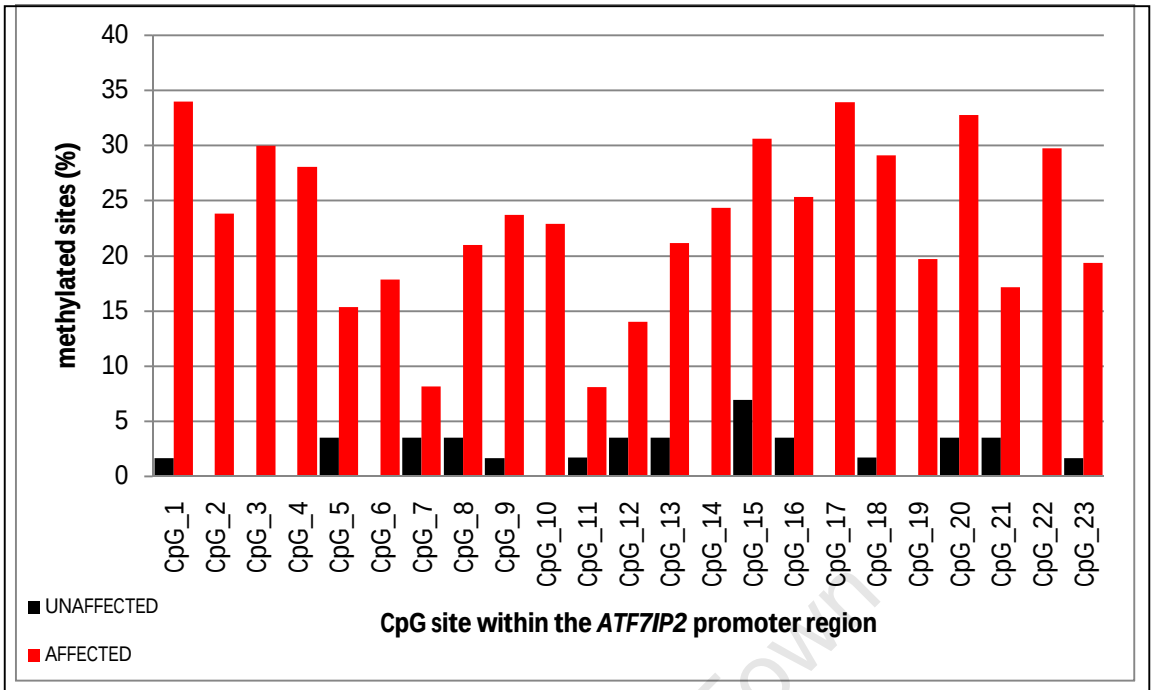


Figure 6.8 The semi-quantitative DNA methylation analysis of the *ATF7IP2* promoter region in *ATRX* mutation positive patients of family XMR2 as well as the unaffected family members. For each of the 23 CpG sites located within this promoter region (x-axis) the corresponding proportion of cells with *ATF7IP2* methylation (y-axis) for the affected (red) and unaffected (black) family members is shown. At all CpG sites the level of methylation was substantially higher (up to 34%) in *ATRX* mutation positive individuals, indicating hypermethylation of this *ATF7IP2* promoter region.

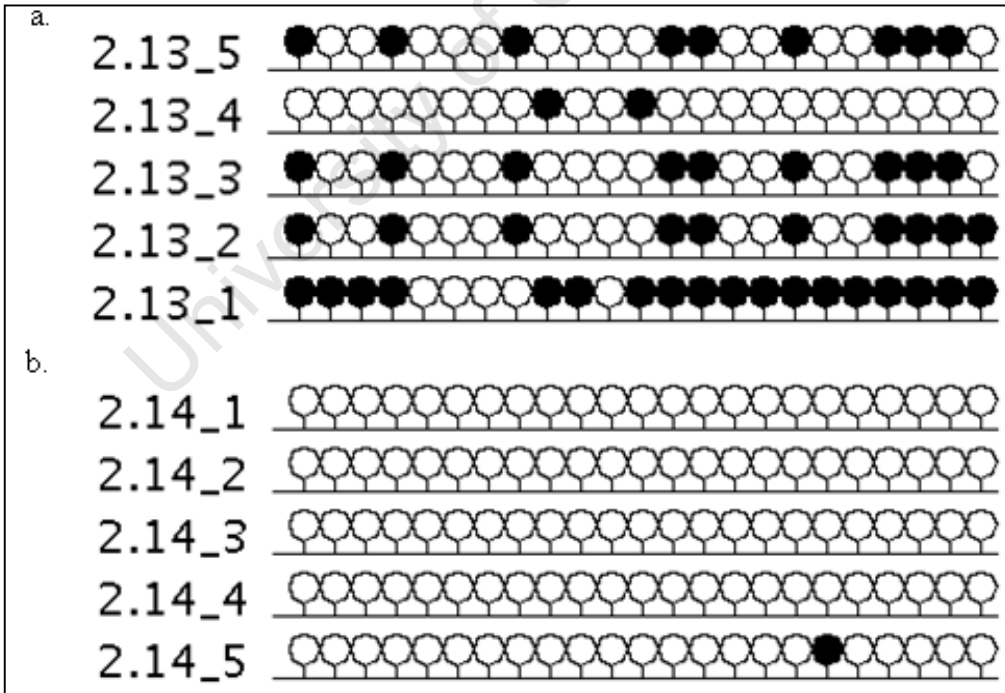


Figure 6.9 'Lollipop' diagrams exemplifying the methylation levels at the *ATF7IP2* promoter region in (a) ATR-X patient (XMR2.13) and (b) unaffected individual (XMR2.14). Each 'lollipop' corresponds to a CpG site within the *ATF7IP2* region, with clear 'lollipops' indicated unmethylated cytosine residues and black 'lollipops' indicating methylation at these sites. It is evident from these DNA methylation representations that the proportion of cells with methylation at this locus in the ATR-X patient supercede those in the unaffected individual at each of these five DNA sequences and is indicative of the hypermethylation at this locus.

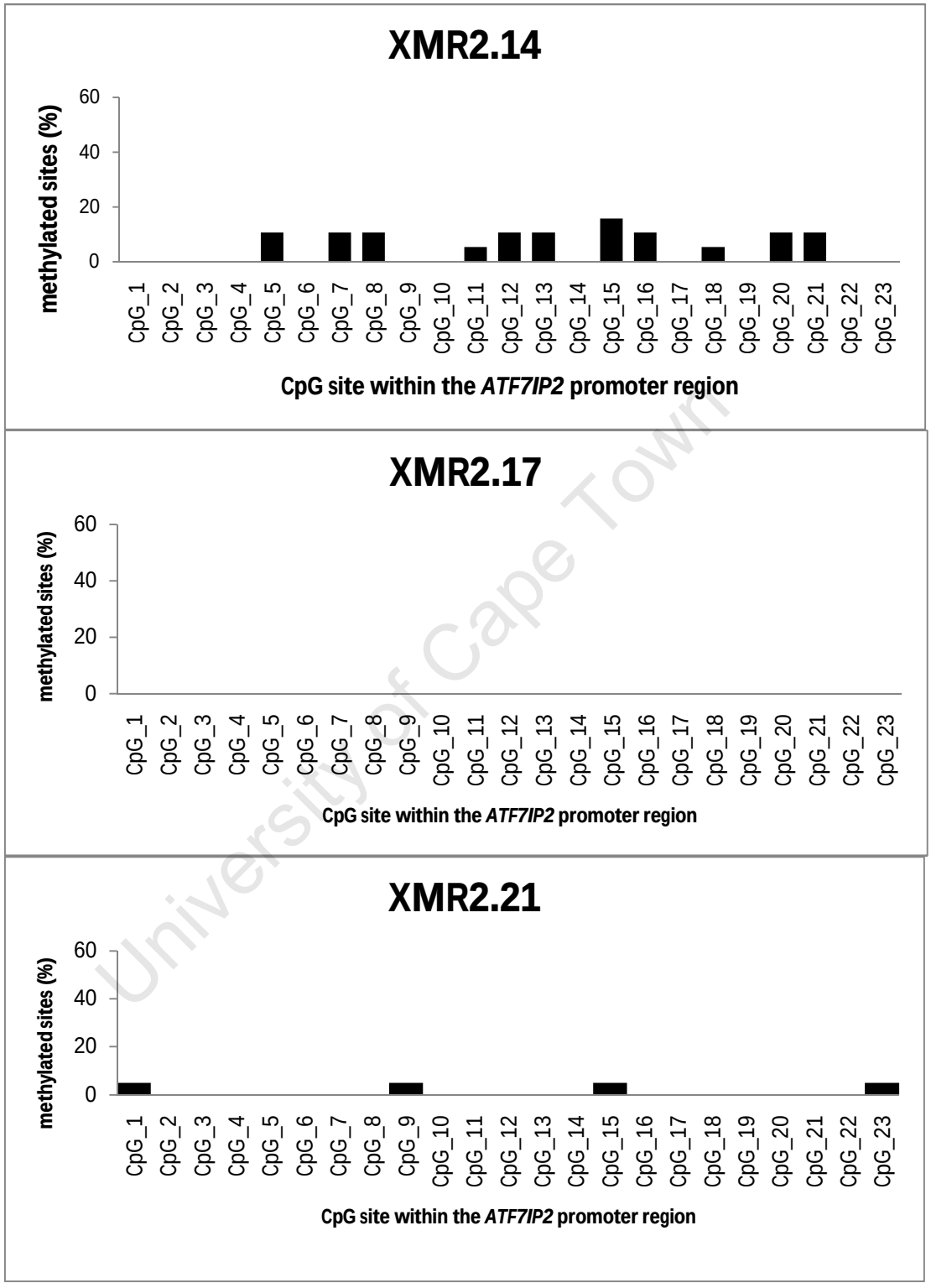


Figure 6.10 The semi-quantitative analysis illustrating the methylation level distribution in the three unaffected family members of XMR2. For each of the 23 CpG sites located within this promoter region (x-axis) the corresponding proportion of cells with *ATF7IP2* methylation (y-axis). Intra and inter-individual variability is evident from these analyses.

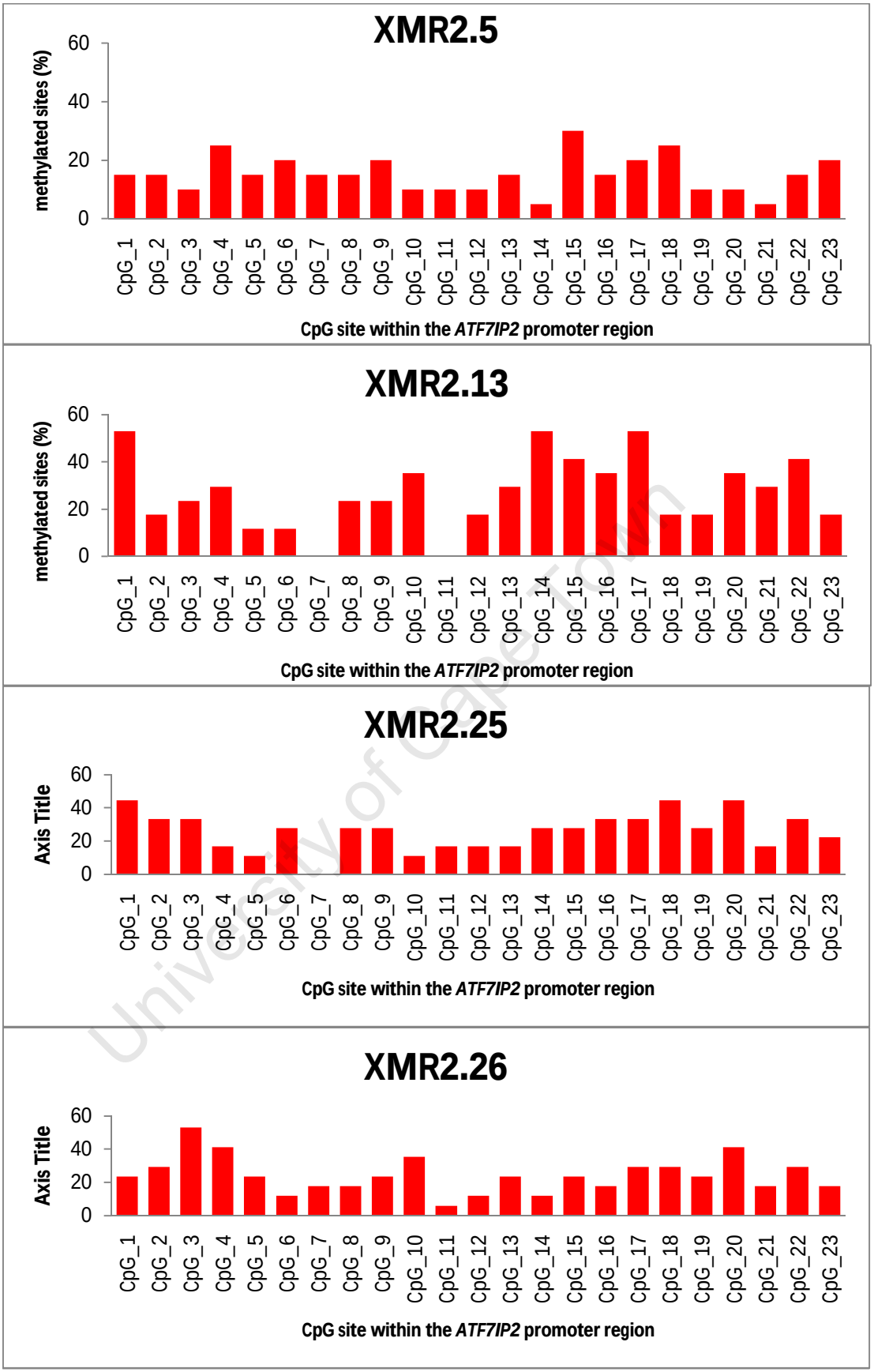


Figure 6.11 The semi-quantitative analysis illustrating the methylation level distribution in the four affected sibs of family XMR2. For each of the 23 CpG sites located within this promoter region (x-axis) the corresponding proportion of cells with *ATF7IP2* methylation (y-axis). Intra and inter-individual variability is evident from these analyses.

6.4 DISCUSSION

The comparative whole-genome DNA methylation analysis conducted between the four ATR-X patients and the three unaffected XMR2 family members identified a total of 112 putative DMRs. Subseqeunt to exclusion of those regions overlapping with know CNVs or SDs, 107 putative DMRs remained of which eight were subject to validation by bisulphite DNA sequencing. These validation experiments demonstrated *ATF7IP2* promoter hypermethylation in ATR-X patients as compared to the unaffected family members of XMR2.

From the total of 112 potential DMRs, five regions (~4%) were excluded on the basis of the DNA sequence overlapping with known copy number variations (CNVs) or SDs. These genomic regions are prone to rearrangements and/or are repetitive in nature; this makes sequence variation at these loci between individuals common. Furthermore, changes in copy number are indistinguishable from methylation enrichment peaks in MeDIP-Chip experiments (A.Sharp, unpublished data). Therefore, it was likely that the observed locus-specific changes in DNA methylation were due to variation in DNA sequence rather than an epigenetic modification thereof (i.e. cytosine methylation). Thus in order to reduce the number of false positive results, these regions were excluded from any further analyses.

From the remaining 107 potential DMRs, the eight DMRs with the highest (nominal and average) absolute t-score values were selected for validation. Of these eight candidate DMRs, one locus, encompassing the *ATF7IP2* promoter region, was confirmed by bisulphite-modified DNA sequencing. The bisulphite DNA sequencing electropherograms generated from each of the four affected individuals demonstrated the presence of both unmethylated (T residues) and methylated cytosines (C residues) at each of the 23 promoter CpG sites investigated. In contrast, for each of the three unaffected family members there was no evidence of methylation at any of these 23 CpG sites as indicated by the occurrence of only thymine residues (unmethylated cytosines). These results illustrated an increase in methylation at the *ATF7IP2* promoter region in the ATR-X patients of family XMR2.

Not surprisingly, the true positive *ATF7IP2* locus had the highest t-score statistic (9.1126, adjusted FDR p-value 2.49E-18) at a single probe, as well as the highest average t-score value (5.367544, adjusted FDR p-value 2.66E-12) across the region. The values of the remaining seven candidate regions had markedly lower t-score statistics (Table 6.1). Also, these results suggest that a significantly high t-score value at a single probe should be accompanied by significantly high t-scores at flanking probes (therefore generating a high average t-score). While an attempt was made to account for the effects of a single deviant probe by selecting DMRs with at least two flanking probes with nominal p-values ≥ 0.01, these results suggest that future studies of this nature should employ stricter DMR selection criteria. In addition, the number of potential DMRs could be reduced by basing DMR selection criteria on a relatively high t-score as well p-value (as appose to just the p-value). However, in defence of the strategy employed here, DNA methylation studies of this nature have not been performed before and therefore a relatively liberal statistical approach was originally employed in attempt to reduce false negative rates. Finally, as with the CNV SNP array analysis (Chapter 2), the importance of validating array results has once again been highlighted.

It was not possible to accurately quantify the proportion of cells with methylation at the *ATF7IP2* locus from the electropherograms generated by bisulphite DNA sequencing. Therefore, a PCR cloning technique was adopted in order to obtain a semi-quantitative measure of methylation ratios in these ATR-X patients. The semi-quantitative analysis of the *ATF7IP2* promoter region using a cloning and colony-PCR technique showed variable methylation levels across the 23 CpG sites of this region. In ATR-X patients the proportion of CpG sites methylated was on average between 8 and 34%, conversely in unaffected family members this averaged methylation level was between 0 and 7%. This result clearly demonstrated the significant degree of *ATF7IP2* hypermethylation as a result of the *ATRX* mutation in family XMR2. Furthermore, this semi-quantitative technique demonstrated the intrafamilial variability of methylation levels at this locus. For instance, there was a small level of methylation of CpG sites in the unaffected family members. Unaffected individual (XMR2.14) exhibited methylation at several CpG sites in up to 8% of cells, conversely individual (XMR2.17) did not demonstrate DNA methylation at any CpG sites for each of the 20 positive clones sequenced. These results suggested that there may be a low level of variable DNA methylation is present in individuals without *ATRX* mutations at the *ATF7IP2* promoter region. Furthermore,

this intrafamilial variability of methylation ratios at this locus was also apparent in the ATRX mutation positive patients, it is likely that this variability is merely similar to the variability observed in unaffected family members. However, it would be pertinent to explore the possibility that residual levels of *ATRX* transcript correlated with the degree if DNA methylation at the *ATF7IP2* promoter region in each affected individual.

Given that hypermethylation at promoter regions leads to a reduction in transcription at the corresponding locus, it is likely that the DNA mehtlyation alteration described in this study leads to a reduction in *ATF7IP2* transcript levels. This deduction could not be tested at this stage due to a lack of mRNA sample availability from XMR2 family members. Also, verification of *ATF7IP2* hypermethylation (and investigation of corresponding transcript levels) in additional ATR-X patients should be performed in order to substantiate the findings described here.

The *ATF7IP2* gene, also known as *MBD1-containing chromatin-associated factor2* (*MCAF2*), is located at 16p13.2. *ATF7IP2* transcribes a 2046bp mRNA that encodes a 681 amino acid protein of approximately 100 kDa (Ichimura et al. 2005) that is known to be expressed in the brain. ATF7IP2 is known to bind to the transcription repression domain of the methylated-cytosine binding protein, MBD1 (Ichimura et al. 2005). Furthermore, ATF7IP2 has also been shown to interact with the transcription activator Sp1 and to a lesser extent (in the cell line investigated i.e. K562), the histone H3K9 methyltransferase, SETDB1 (Ichimura et al. 2005). The overlapping protein interactions of ATRX and ATF7IP2 suggest that these two proteins may form part of the same protein complex or repressive chromatin assembly pathway. This hypothesis is supported by several lines of empirical evidence (Figure 6.12). Firstly, ATF7IP2 forms a complex with SETDB1, responsible for the methylation of H3K9 (Ichimura et al. 2005). In turn, this $H3K9_{me3}$ mark has been shown to elicit ATRX localisation at Y-specific repeats and pericentric heterochromatin as well as to the Xist locus during X-inactivation (Baumann et al. 2008, Baumann and De La Fuente. 2009). Also, ATR-X patients exhibit marked hypermethylation at Y-specific repeats (Gibbons et al. 2000). In addition, $H3K9_{me3}$ serves as a mark for HP1, this protein localises with ATRX at heterochromatic regions (McDowell et al. 1999). These data suggest that ATRX targeting to these heterochromatic regions, which is dependent on $H3K9_{me3}$, may be reliant on the histone methyltransferase activity of SETDB1 through co-operation with ATF7IP2.

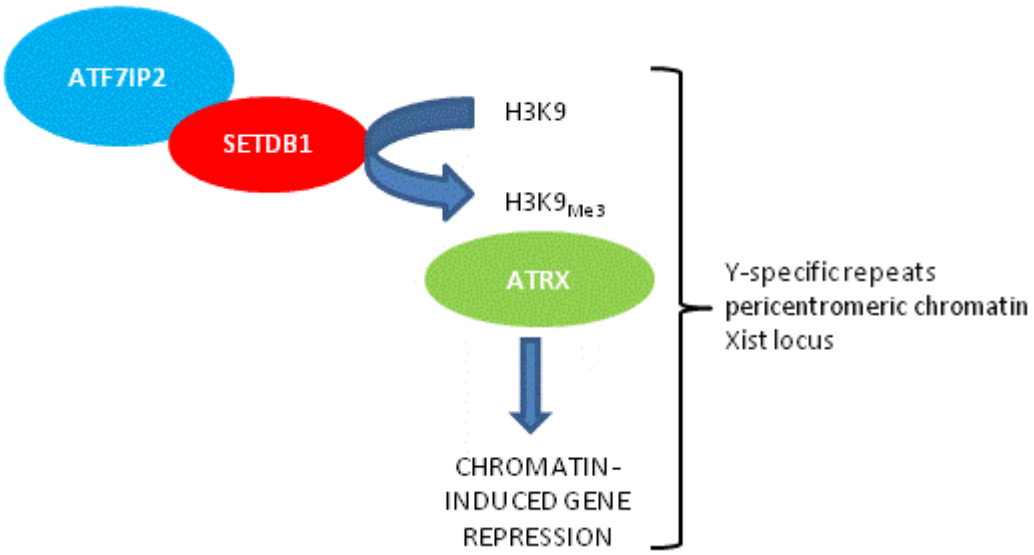


Figure 6.12 The theoretical complex involving ATRX and ATF7IP2. ATF7IP2 controls the histone methyltransferase activity of SETDB1. H3K9me3 at specific loci leads to the recruitment of ATRX and subsequent chromatin silencing

ATF7IP2 has also been shown to bind MBD1 (Ichimura et al. 2005), moreover, methylated cytosine residues and the subsequent recruitment of MBD proteins is a mark of repressive chromatin. Interestingly, ATRX also interacts with another MBD protein, MeCP2 (Nan et al. 2007), further suggesting a similar repressive chromatin remodelling function for ATRX and ATF7IP2. It would therefore be interesting to assess the *ATF7IP2* promoter DNA methylation status in *MECP2* mutation positive patients, in order to test the extent of a putative overlapping function between ATRX and ATF7IP2, as well as the respective associated MBD proteins.

There is also evidence that ATF7IP2 may have a transcription activating role in addition to the aforementioned repressive function. ATF7IP2 associates with the transcription activator Sp1 (Ichimura et al. 2005), known to localise to the PML-NBs, sites of active transcription (Vallian, Chin and Chang. 1998). Furthermore, ATRX has been shown to interact the with transcription co-activator DAXX at the PML-NBs (Xue et al. 2003). Once again these overlapping transcriptional regulatory roles of ATRX and ATF7IP2 suggest these proteins may participate in similar gene regulation pathways. While the aforementioned empirical data provides evidence suggesting ATF7IP2 and ATRX have similar repressive/activating chromatin remodelling properties and potentially function in the same protein complex, this hypothesis requires further experimental validation. Such experiments could take the form of yeast two-hybrid assays or immunoprecipitation studies.

It remains to be established whether the ATF7IP2 promoter hypermethylation in ATR-X patients is due to direct targeting by ATRX. Alternatively, assuming hypermethylation leads to reduced *ATF7IP2* transcript levels, it is tempting to speculate that these proteins form part of a transcription remodelling complex which requires proportional levels of ATRX and ATF7IP2. It has been previously suggested that ATRX is required in precise stoichiometric amounts for protein function (Berube et al. 2002, Picketts et al. 1996). Hence, in the presence of reduced ATRX levels, ATF7IP2 production would be concomitantly reduced by DNA methylation of the *ATF7IP2* promoter region. It would therefore be interesting to assess whether residual levels of *ATRX* transcripts correlated with a putative reduced *ATF7IP2* mRNA levels in ATR-X patients.

Alterations to the *ATF7IP2* transcript have not been previously associated with any disease phenotype. Furthermore, no MR chromosomal aberrations or MR candidate loci have been reportedly linked to this region. However, the overlapping function of ATF7IP2 and ATRX suggests that *ATF7IP2* is a good candidate MR gene. With the expected increase in CNV testing in MR patients, as well as homozygosity mapping in autosomal recessive families, it is expected that the number of autosomal MR loci will concomitantly increase. In such instances, ATF7IP2 would be a mutation screening candidate in patients linked to this locus. Therefore, in this study a novel strategy for the identification of MR candidate genes has been presented. Furthermore, using this novel research approach it has been shown that alterations to the DNA methylation profile is evident in MR patients. Application of this technique to mutation negative MR patients may provide novel insights into the role of this epigenetic mechanism in MR pathogenesis and furthermore identify new candidate MR genes.

The results presented here show that *ATRX* aberrations do not lead to wide-spread changes at CpG promoter regions. The array design selected for this study (Whole-Genome CpG island and promoter array) targets only the genomic regions which are associated with CpG sites at promoter regions. Therefore, it does not cover intragenic and intergenic regions which can also be subject to cytosine methylation, though modification of these regions is less well-studied. Previously it was shown that *ATRX* mutation positive patients have DNA methylation aberrations at specific loci which do not correspond to promoter regions. These loci included the rDNA gene arrays which localise to the short arms of acrocentric chromosomes (chromosomes 13, 14, 15, 21 and

22). These regions were largely unmethylated in patients as compared to normal individuals (up to 20% of rDNA CpG sites were methylated) (Gibbons et al. 2000). The whole-genome CpG promoter HD2 array (Nimblegen) used in this study does not have probe coverage at these acrocentric regions, most likely due to the repetitive nature of the satellite tandem repeats which comprise these regions. Similarly, the hypermethylation of Y-specific repeats (DYZ2) in ATR-X patients observed by Gibbons and colleagues (2000) was not replicated in this study due to a lack of probe coverage in these repetitive regions (Gibbons et al. 2000). However, given that no genome-wide DNA methylation studies in ATR-X patients have been conducted to date and that DNA methylation is most well-studied at CpG promoter regions, this array platform was selected for these investigations.

The DNA samples selected for these investigations were isolated from lymphocytes; this approach may also account for the lack of multiple aberrantly DNA methylated sites in ATR-X patients. Therefore, these experiments would ideally be performed in the primary affected organ, the brain. However, this tissue is difficult to obtain and requires the acquisition of post-mortem patient samples. Alternatively, similar DNA methylation investigations could be performed in the conditional *ATRX* knock-out models such as those generated by Berube and colleagues (2005) (Berube et al. 2005), or using siRNA ATRX knockdown in human neuronal cell lines.

The whole-genome promoter DNA methylation approach adopted in this study has been successfully employed for the detection of an aberrantly methylated gene, *ATF7IP2*. The results of these investigations have therefore provided a novel research direction for further investigations of a potential repressive chromatin complex including ATRX and ATF7IP2. These future studies will contribute to the understanding of the pathogenesis of ATR-X and related syndromes. Furthermore, *ATF7IP2* would present a good candidate for mutation screening in autosomal MR patient cohorts that, by future homozygosity mapping studies, or CNV studies, are shown to be linked to the 16p13.2 locus. Finally, while the effects of ATRX ablation on promoter DNA methylation are not widespread, these investigations show that changes to cytosine methylation are a molecular feature of MR in at least one XLMR gene. Employing the methodology presented in this study for DNA methylation aberration investigations in additional XLMR 'epigenetic' genes, will establish the extent of the role of this epigenetic mechanism in the pathogenesis of this disorder.

CHAPTER 7

FINAL DISCUSSION AND CONCLUSIONS

MR is a debilitating condition affecting a significant proportion of the population. While the causes of MR are extremely heterogeneous and include both genetic and environmental factors, recent years have seen major advances in defining the genetic basis of this disorder. These advances can primarily be attributed to the rapid expansion of molecular genetic techniques, in particular high-throughput DNA sequencing and microarray technologies. By implementation of these techniques, a total of 87 XLMR genes have been identified to date. Of these genes, 45 (52%) have been discovered in the last five years, with the remainder being identified in the preceding 13 years (Chiurazzi et al. 2008, Cho et al. 2008, Molinari et al. 2008, Tarpey et al. 2009). However, despite this exponential growth in the number of XLMR genes, up to 50% of XLMR remains unexplained (de Brouwer et al. 2007, Gecz, Shoubridge and Corbett. 2009).

It is reasonable to conclude that the remaining unexplained XLMR cases can be attributed to novel XLMR genes. Also, while the role of chromosomal aberrations in MR has been well-established, the more recent application of microarrays to detect even smaller CNVs has led to an increase in the number of disease-causing copy-number changes. It is expected that as the use of microarrays for CNV detection becomes more widespread, a significant proportion of unexplained MR cases will be resolved by CNV identification. However, it is also expected that novel disease mechanisms, including mutations in non-coding regions or epigenetic changes, will be discovered. In lieu of this supposition, the role of DNA methylation in XLMR pathogenesis has been alluded to (Froyen et al. 2006), but not investigated. Furthermore, existing molecular techniques need to be adapted to detect changes in the DNA methylation profiles in affected individuals. To this end, this study tested the hypothesis that specific XLMR genes exert their pathogenic effects by altering the DNA methylation profile of mutation positive individuals.

The research strategy adopted in this study comprised of two main components. Firstly, in Part A, a variety of molecular techniques were employed to detect mutations in XLMR 'epigenetic' genes. This aspect of the study included: CNV analysis, linkage analysis and positional candidate gene mutation screening, as well as functional candidate gene mutation screening. From these analyses, those individuals whose etiology was defined by a disease-causing mutation were prioritised for DNA methylation profiling in the secondary phase of this dissertation. Ultimately, in Part B, the ATR-X family identified in this study was selected for whole-genome DNA methylation profiling. The elucidation of DNA methylation alterations in these patients illustrate that changes to this epigenetic mechanism are indeed apparent in XLMR patients, if only on a limited scale, for at least one XLMR gene. However, collectively these molecular investigations address XLMR pathogenesis on a much wider scale. In the sections that follow, the findings of this study are discussed in relation to three main themes: XLMR gene identification, unravelling the pathogenic role of epigenetic molecular mechanisms, and finally a brief look towards the future therapeutics which may ultimately be developed for patients that suffer from this debilitating condition.

7.1 XLMR: IDENTIFYING A GENETIC CAUSE

7.1.1. CONTRIBUTION OF DISEASE-CAUSING CNVS

Since the application of microarray technologies to the identification of MR-causing CNVs numerous studies have reported high detection rates of up to 24% (Aston et al. 2008) (Table5.1). Therefore, in collaboration with colleagues at the Department of Human Genetics, Nijmegen, The Netherlands, 30 South African MR patients presenting with MR and additional clinical manifestations were assessed for disease-causing CNVs using a 250K SNP array. These investigations led to the detection of three disease-causing alterations, 17p11.2Del, Xq25Dup and lastly the Xq26.3-27.3Del which exhibits low penetrance. In addition, two CNVs of indeterminate disease-causing nature, 6q24.3 and 4p16.3Del, were detected.

The approximately 3Mb 17p11.2Del was detected in individual Fx343.1. This CNV was shown to overlap with the known Smith Magenis microdeletion syndrome chromosomal region. Therefore, the MR and associated phenotype in this individual was concluded to be due to the Smith Magenis Syndrome, which arose from a de novo

mutation. Furthermore, a retrospective review of the patient's clinical records (which were not available to us prior to this study) indeed revealed that this patient had been previously diagnosed with Smith Magenis syndrome by FISH.

The second disease-causing CNV detected in this study was the de novo Xq25Dup detected in individual XMR8.1. The Xq25Dup encompasses 4 known genes (*GRIA3, THOC2, XIAP, and STAG2*) of which, *GRIA3* is a known XLMR gene (Wu et al. 2007). Duplications in this chromosomal region have been previously reported (Bonnet et al. 2009, Chiyonobu et al. 2007). The duplications described in these previous reports encompass *GRIA3* (Chiyonobu et al. 2007) and the *GRIA3*, *XIAP*, *STAG2* genes (Bonnet et al. 2009). Thus, in contrast to the duplication described in this study, neither of these previously reported duplications encompassed the *THOC2* gene. Given the more severe presentation of individual XMR8.1 as compared to the patients described in previous reports, it is hypothesised here that *THOC2* may play a role in the pathogenesis of MR in this individual. These results highlight the potential that CNV detection holds for the identification of novel MR genes. The delineation of the dosage-sensitive genes within disease-causing CNVs presents a novel way to identify candidate genes for mutation detection in XLMR/MR cohorts.

The 10Mb Xq26.3-27.3Del detected in family Fx444 exhibited variable phenotypic expression. Family Fx444 presented with a typical X-linked inheritance; the proband presented with severe MR, haemophilia B, Lennox Gestault epilepsy syndrome and multiple dysmorphic features. The proband's deceased maternal cousin presented with a similar phenotype, while his half-brother presents only with Haemophillia B and mild learning disability. The half-brothers and the carrier mother were all positive for the Xq26.3-27.3Del. The region encompasses 21 genes, including *F9* (mutations in which cause Haemophillia B) and *SOX3*, previously shown to be mutated in MR patients. The remaining 19 genes encode for proteins whose function do not make them candidates for pathogenesis of this disorder, or else they constitute part of proteins which may have redundant function. Therefore, due to the phenotypic variability observed between patients with *SOX3* mutations (Solomon et al. 2004, Woods et al. 2005), as well as *Sox3* knock-out mice (Rizzoti et al. 2004), it was concluded that the Xq26.3-27.3Del causes the syndromic MR presentation in family Fx444, but that this deletion exhibits low penetrance.

The detection of this low penetrance Xq26.3-27.3Del in family Fx444 also highlights the care that should be exercised when interpreting the disease-causing nature of detected CNVs. In this instance, it is particularly difficult to anticipate the phenotypic outcome of males carrying this deletion, complicating genetic counselling and future management of family Fx444. Furthermore, similar CNVs with variable penetrance have been reported. A 16p11.2Del encompassing an exact genomic interval was identified in individuals with autism, MR with/without MCA, speech and learning problems, as well as normal cognitive function (Bijlsma et al. 2009). Similarly, the 15q13 microdeletion/duplication has been associated with inter and intra-individual variability resulting in a spectrum of phenotypes, from severe MR with additional features, to complete lack of a recognisable disease phenotype (van Bon et al. 2009). As the use of whole-genome arrays becomes more wide-spread it is anticipated that additional cases of low penetrance and phenotypic variability, such as those aforementioned, will be reported. It will be important to develop international, as well as institutional guidelines for the delivery of such reports and subsequent genetic counselling of families.

In the CNV analysis described in this study, the pathogenic nature of two variants, 6q24.3 in Fx151.1 and 4p16.3Del in XMR14.1, could not be determined due to the unavailability of parent DNA to assess the segregation of these variants in the families. As in previous studies, these results highlight the importance of the availability of parents for testing in order to attribute a CNV molecular MR diagnosis. Thus, taken together, parent availability will be an important consideration during the development of protocols and procedures for the implementation of CNV testing in molecular diagnostics in South Africa.

As CNV studies become more widespread and as this technology is increasingly incorporated as a first-line diagnostic tool, it is concomitantly expected that the number of microdeletion/duplication syndromes will increase. In order to facilitate molecular diagnosis of these syndromes it is critical that these disorders are catalogued in freely available, web-based databases. The database of chromosomal imbalance and phenotype using Ensembl resources (DECIPHER) (https://decipher.sanger.ac.uk/application/) is an example of such a database (Firth et al. 2009). DECIPHER, based at the Wellcome Trust Sanger Institute, Cambridge, UK, catalogues known and novel disease-causing CNVs and associated phenotypes from

participating centres world-wide. In addition, the CNP compilation allows easier dissection of CNVs detected in array analyses. With the accumulation of information pertaining to genomic architecture, and its role in MR, databases such as these will be critical in MR diagnosis due to CNVs.

The CNV detection rate in this study was 1/17 (6%) in XLMR families (assuming the Xq26.3-27.3Del in family Fx444 was indeed the disease-causing mutation). However, this detection rate should be interpreted with caution given the small sample size, even though the rate was consistent with that reported in previous studies (Lugtenberg et al. 2006, Madrigal et al. 2007). Despite, the numerous reports of X-chromosomal CNVs accounting for XLMR, to date only two studies have analysed copy number in XLMR cohorts exclusively. Lugtenberg et al. (2006) screened 40 NS-XLMR families using a whole X-chromosome screen and identified three disease-causing mutations, hence resolving 7.5% of XLMR cases studied (Lugtenberg et al. 2006). Madrigal and colleagues (2007) screened 52 XLMR patients using an X-specific BAC array and detected five disease-causing mutations (pick-up rate 9.5%) (Madrigal et al. 2007). Together with the results of this dissertation, these studies strongly motivate for systematic screening of XLMR families using commercially available whole X-chromosome specific arrays, or alternatively, X-chromosome specific Multiplex Amplifiable Probe Hybridisation (MAPH) (Kousoulidou et al. 2007), or exon-specific X-chromosome arrays (Bashiardes et al. 2009).

In addition to the Xq26.3-27.3Del identified in XLMR family Fx444, this study led to the detection of two additional disease-causing mutations, (Xq25Dup in XMR8, and 17p11.2 in Fx343) in the total of thirty MR patients assessed (10% pick-up rate). Furthermore, CNVs have been demonstrated to account for up to 24% of MR in some cohorts (Aston et al. 2008) (Table 5.1). Collectively this evidence suggests that CNV analysis using whole-genome arrays in isolated/autosomal dominant MR cases as well as X-specific arrays in XLMR families would be a critical addition to the repertoire of MR diagnostic tests offered by the NHLS. However, at present, the department lacks the technologies required to perform microarray experiments (hence this study was conducted at the Department of Human Genetics, Nijmegen Medical Centre, The Netherlands). Therefore, until the necessary infrastructure is available (funding motivations will be supported by this study), CNV testing will be focused on known

microdeletion syndromes and subtelomeric regions using MLPA techniques, under the auspices of the NHLS.

The establishment of a genomics platform at UCT for CNV detection is critical not only for MR diagnostics but also for the advancement of medical genomics research on a broader scale. The importance of CNVs in the pathogenesis of several disorders including schizophrenia (Walsh et al. 2008), autism (Miller et al. 2009), HIV susceptibility (Gonzalez et al. 2005) and Crohn disease (Fellermann et al. 2006), have been demonstrated recently. It is expected that in the future numerous additional disorders will found to be explained, at least in part, by mutations that affect copy number of the human genome.

7.1.2. XLMR MUTATIONS IN CODING REGIONS

7.1.2.1 Linkage analysis and positional candidate gene screening

Linkage analysis and subsequent positional candidate gene screening has been successfully implemented in the identification of numerous genes for both XLMR and a multitude of additional disorders. In this study, this research strategy was successfully adopted to identify the *ATRX* disease-causing mutation in family XMR2. Initial whole-X chromosomal linkage analysis in 12 individuals of family XMR2 showed linkage of XLMR to a 38Mb critical interval. Additional fine-mapping in the region, as well as genotyping of four additional family members led to a reduction of the critical chromosomal interval to a 10Mb region at Xq13.1-q21.1. Encompassed within this linked interval were nine known XLMR genes. Based on phenotypic expression of known *ATRX* mutation positive patients, this gene was selected as the primary candidate gene in family XMR2. These results illustrate the importance of good clinical characterisation of affected family members in order to prioritise candidate genes.

By mutation screening of the three *ATRX* mutation hotspots in family XMR2 a c.5987_6011del mutation was detected in exon 26. This 24bp deletion was predicted in silico to give rise to a truncating mutation (p.1987fsX56). However, aberrant splicing, due to disruption of an ESE site, results in the deletion of 66bp from the mRNA (r.5957_6022del) and ultimately an interrupted protein (p.1987_2008del). These results illustrate two interesting observations; firstly, the role that ESE sites play as a disease

mechanism and secondly, the importance of phenotypic rescue observed for *ATRX* mutations.

There are now numerous reports of ESE site disruption resulting in aberrant splicing of pre-mRNA transcripts and ultimately human disease (Boichard et al. 2008, Collin et al. 2008, Ramser et al. 2005). Of particular importance are the reports of synonymous (Collin et al. 2008) and silent (Boichard et al. 2008, Ramser et al. 2005) mutations leading to aberrant splicing, these examples exemplify the necessity to characterise genomic mutations on a mRNA transcript level. There are most-likely a number of variants identified by XLMR gene screening projects which are classified as benign which in actual fact affect splicing. Investigation of these 'benign' variants could be assisted by the use of prediction programmes including ESEFinder used in this study and others as reviewed by Houdayer and colleagues (Houdayer et al. 2008).

The ATR-X patients in family XMR2 exhibit phenotypic rescue by alternative splicing of the *ATRX* transcript. The genomic deletion in this family was predicted to give rise to a truncated protein, however, an alternative splicing event, putatively due to alternative ESE site usage, allows skipping of the truncating event and generates an interrupted full-length protein. This nonsense-associated alternative splicing (NAS) has been described in several other disorders including Duchenne Muscular Dystrophy (DMD) (Nishiyama et al. 2008), pyruvate dehydrogenase (PDH) deficiency (Ridout et al. 2008) and cystic fibrosis (Aznarez et al. 2007). NAS can be induced by a number of either intronic or exonic factors including enhancers or silencers. However, the extent of this phenotypic NAS rescue mechanism in the pathogenesis of XLMR is not known. To date, to the author's knowledge, only two XLMR genes have been shown to make use of this mechanism. Firstly, the *ATRX* mutation as described in this study, as well as all patients with truncating mutations downstream of the ATRX ADD domain, exhibit similar NAS events (Gibbons et al. 2008). Secondly, more recently, mutations in the XLMR gene, *PQBP1*, were shown to elicit NAS by enhanced skipping of the mutated exon 4 in a mutation-specific manner (Musante et al. 2010). Similar investigations in XLMR gene mutation positive individuals will enhance our understanding of NAS in the pathogenesis of XLMR and perhaps provide novel insights for putative therapeutic targets.

The clinical presentation of family XMR2 was consistent with phenotypes described in previous *ATRX* mutation positive patients. Importantly, an affected individual in this family do not present with α-thalassaemia. Therefore, this observation supports the suggestion that this clinical feature should not be a defining criterion for *ATRX* mutation screening (Villard et al. 1999).

In this study, a mutation protocol technique was developed to detect mutations in the three *ATRX* mutation hotspots which account for 76% of all known mutations in this gene. This mutation detection protocol entailed isolation of mRNA from the affected individual, after a cDNA synthesis step, PCR and subsequent DNA sequencing were performed for the three mutation hotspots. The amplification and analysis of just three regions greatly reduced the cost associated with mutation detection within this gene. This protocol has been incorporated into the diagnostic protocols of the NHLS. Given the rarity of the condition, to date, only a single individual has been referred to the Division for diagnostic testing. However, it is envisaged that, in the future, the implementation of this diagnostic test will still have a positive influence on the alleviation of the unexplained MR cases for a small number of families in South Africa.

Linkage analysis in a smaller family, Fx56, resulted in the definition of a much larger chromosomal region (Xp11.4-Xq21.1) spanning 76Mb and encompassing approximately 397 genes. Genotyping of additional markers did not result in reduction of the disease-associated region. Also, no additional family members could be recruited to this study, due to problems re-establishing contact with the family. Therefore, this critical interval could not be refined and it was not possible to prioritise positional candidate genes for mutation screening.

The problems encountered with re-establishing contact with families not only in the linkage phase of this dissertation, but also throughout the course of this study was a limiting factor, which needs to be addressed in future studies. These problems are a reflection of capacity problems that Divisional clinicians experience as well as the overloading of the public health system, both matters need to be addressed on a national level. At a divisional level, some of these deficits could be ameliorated by storage of bloods/plasma at -80°C, or preferably cyropreserved (Louie and King. 1991), thus preserving the integrity of the cells so that patient cell lines (EBV-transformed lymphoblasts) could be established if needs be. In addition, it would be recommended

that future XLMR/MR studies are conducted prospectively rather than retrospectively so as to avoid problems with contacting family members.

Family Fx67 consisted of three affected brothers and their unaffected mother, due to the family's small size, linkage analysis could not be conducted, as a meaningful LOD score could not be generated. However, allele segregation analysis spanning the X chromosome, showed co-segregation of six microsatellite markers with the disorder in this family. No additional family members were available for genotyping and the co-segregating region could not be further reduced in this family. The six microsatellite markers spanned 11Mb of a relatively gene poor region (Xp21.1-Xp11.4) which consists of 27 known genes, of which two (*ATP6AP2* and *TSPAN7*) have been previously implicated in XLMR.

Patients previously reported with mutations in *ATP6AP2* and *TSPAN7* present with NS-XLMR. Concomitantly, patients in family Fx67 also exhibit NS-XLMR; therefore, *ATP6AP2* and *TSPAN7* were subject to mutation screening in the three affected individuals, but no disease-causing mutations were identified. While, it could not be ruled out that the MR presentation in this family was not X-linked, these results are suggestive of Fx67 harbouring a novel XLMR gene mutation. Unfortunately, at this stage, the remaining 25 genes within the putative disease-associated chromosomal interval were not subject to mutation detection in family Fx67. The reasons for this were two-fold; firstly, a significant LOD score could not be generated for this family. Secondly, given the genetic heterogeneity of MR, it was difficult to prioritise candidate gene screening within this interval especially given the NS-XLMR presentation of affected individuals in Fx67.

The difficulty experienced in prioritising candidate genes was evident not only in this study, but has also been well documented though-out the years of XLMR research. Bioinformatics approaches, such as those discussed by Oti and Brunner (2007) cannot be applied in the context of MR research given the vast number of functional pathways in which XLMR genes participate (Oti and Brunner. 2007). In order to circumvent this problem, it is envisaged that mutation screening of multiple candidate genes will become feasible as high-throughput DNA sequencing becomes more accessible and cost-effective.

These linkage analyses and positional candidate gene screens demonstrate that as medical and genetic infrastructure becomes more apparent in less developed countries, it will be possible to conduct molecular investigations in additional XLMR families. Using relatively cost-effective techniques such as those described here, it is likely that mutations will be found in these families in both novel and known XLMR genes. Thus, it was demonstrated here, that studies such as these are important in elucidating novel causes of XLMR, this is particularly important in context of XLMR given the high percentage (>50%) of unexplained MR cases.

7.1.2.2 Screening functional candidate genes

The *ARX* gene

After the CGG expansion mutation in the *FMR1* gene promoter region, *ARX* mutations are the second biggest contributor to XLMR. In this study, 113 clear (43) or putative (70) XLMR families were assessed for mutations in the five *ARX* exons using dHPLC. DNA sequencing of samples with deviant dHPLC profiles revealed two c.428_451dup mutations in Fx391.1 and Fx446.1, the carrier mothers of these individuals were also found to be positive for this mutation. Also, a c.300G>A was identified in Fx362.1; however, this variant did not segregate with the disorder in this male sib-pair. Furthermore, screening for the two common polyalanine *ARX* expansion mutations (c.428_451dup and c.304ins$(GCG)_7$) in 183 isolated male MR patients revealed no alterations in these patients.

The c.428_451dup is the most common disease-causing *ARX* mutation and accounts for 39/84 (46%) of all reported cases (Appendix 5A). The identification of the two c.428_451dup expansions in this study increase the preponderance of this second polyalanine tract expansion in *ARX* mutation positive patients to 48%. The second most common *ARX* mutation is an expansion of the first polyalanine tract, mainly the c.304ins$(GCG)_7$, which accounts for an additional 15% of all mutations described to date. Therefore, collectively, the polyalanine expansion mutations account for 63% of all known *ARX* mutations.

In individual Fx362.1 a c.300G>A mutation was detected, this mutation did not segregate with the disorder in the family as it was not detected in this individual's affected brother, Fx362.2. Therefore, this alteration was not the direct disease-causing

mutation in this family. However, it was interesting to note that the c.300G>A mutation was predicted to alter an ESE site, Srp55; similar ESE site disruptions have been reported in *ARX*. Gronskov and colleagues (2004) described an *ARX* c.1347C>T variant, that similarly affected a Srp55 ESE site, in un/affected XLMR patients and speculated that this variant could act as a modifier (Gronskov et al. 2004). Second to this, through a collaborative study, an *ARSF* truncating variant (p.S76fsX7) was identified in individual Fx362.2 but not Fx362.1 (Tarpey et al. 2009). Thus, while neither of these variants are the disease-causing mutation in this brother-pair it is speculated here that an accumulation of similar protein truncating or modifying variants could lead to XLMR in family Fx362 in the mould of a multifactorial of "x-linked risk factor' hypothesis. In support of this hypothesis, in the study by Tarpey et al. (2009), truncating variants were present in up to 1% of X-linked genes in the general population (Tarpey et al. 2009), suggesting that there may be a threshold above which the accumulation of these variants manifests in XLMR, specifically non-syndromic, mild MR.

The pick-up rate for *ARX* mutations in this study was 2/43 (4.5%) among clear XLMR families. Despite this low mutation rate (4.5%) as compared to previous international reports of detection rates spanning 6.6-9.5% (de Brouwer et al. 2007, Mandel and Chelly. 2004, Poirier et al. 2006), this figure was significant enough to warrant the development of an *ARX* diagnostic test. Given the cost-effectiveness of the PCR-based approach, the screening of the common *ARX* mutations (c.428_451dup and c.304ins$(GCG)_7$) for XLMR families was implemented as part of the diagnostic testing protocols of the NHLS in late 2007. Since its inception a patient with a c.304ins$(GCG)_7$ polyalanine *ARX* expansion has been identified from a total of 19 referred patients. Thus, the incorporation of this diagnostic test has already had a positive impact on alleviating the burden of unexplained MR in the Western Cape region of South Africa and will continue to do so in the future.

The screening of 183 isolated male MR cases for the two most common ARX polyalanine expansion mutations revealed no mutations in this group of patients. This zero detection rate for isolated male MR cases was in keeping with previous international reports; Gronskov et al. (2004) reported a 0.3% detection rate while de Brouwer et al. (2007) reported a pick up rate of 0.1% (de Brouwer et al. 2007,

Gronskov et al. 2004). Collectively, these detection rates indicate that *ARX* polyalanine expansion diagnostic testing is not advised in isolated cases of MR.

The *KDM5C* gene

After, *FMR1* and *ARX*, *KDM5C* is one of the more commonly mutated XLMR genes. Thus, in this study 25 patients presenting with clinical features found to be predominant in *KDM5C* mutation positive patients were assessed for mutations in the 26 *KDM5C* exons using dHPLC. Three alterations, c.351-38C>T in Fx135.1, c.2623-45G>A in Fx277.1 and c.2517-7_8insTAC in Fx361.1, were detected by DNA sequencing of these samples which exhibited altered dHPLC profiles.

The c.351-38C>T, originally detected in individual Fx135.1, was detected in three out of 100 ethnically matched controls using a PCR and RE digest approach. Therefore, given that this variant was not previously reported in web-based databases (dbSNP at NCBI or Ensembl), it was concluded that the c.351-38C>T variant constitutes a novel SNP. Thus, this variant was not the disease-causing variant in family Fx135.

The c.2623-45G>A was originally identified in Fx277.1; this variant had not been previously described. It was not possible to screen for this variant using either RE digest or ARMS-PCR in the background population as neither of these techniques was successfully optimised. While more expensive technologies are available to screen for this variant, these techniques were not implemented due to the deeply intronic nature of this alteration and thus the minimal likelihood that it contributes to disease in this family.

The c.2517-7_8insTAC was detected in Fx361.1. in this study. This variant has been previously described in a family with XLMR, but was not detected in 312 control X chromosomes screened (Jensen et al. 2005). In concordance with these previously published results, the c.2517-7_8insTAC was not detected in 100 unaffected males using an ARMs-PCR screening technique in this study. Thus, cumulatively, this variant has been detected in two XLMR patients to date, but not 412 control X chromosomes. In addition, the c.2517-7_8insTAC lies directly adjacent to the pyrimidine-rich branch site which is essential to correct pre-mRNA splicing. Collectively, this evidence suggests that this alteration is the disease-causing mutation in these XLMR families, but verification by mRNA sequencing in either of these patients is required to definitely confirm this supposition.

Of the total of 501 XLMR families and sib-pairs investigated in large cohorts internationally, a total of 18 (3.2%) *KDM5C* disease-causing mutations have been described (de Brouwer et al. 2007, Santos et al. 2006). Thus it was not unexpected that no *KDM5C* mutations were detected in this study, despite the fact that this South African study cohort was enriched based on a clinical presentation previously described in mutation positive patients. An investigation in even larger, clinically relevant South African patient cohorts will give a better approximation of the incidence of *KDM5C* mutations in this population group. It is anticipated that these studies will be aided by future developments in next generation DNA sequencing technologies, both in terms of cost and accessibility.

Future functional candidate gene screening studies

Additional mutation screening investigations in XLMR genes with high mutation frequencies, such as the *ARX* and *KDM5C* investigations described here, will guide the development of additional MR diagnostic tests in South Africa. Amongst the more commonly mutated genes, *MECP2* duplications are one of the most prevalent. To this end, the incidence of *MECP2* duplications are being investigated within this research group at the Division. Preliminary results suggest successful mutation detection, which further substantiates the feasibility of diagnostic *MECP2* duplication testing for XLMR patients in the future.

With regards to future mutation screening in functional candidate genes it is critical that detailed clinical information is obtained from clinicians referring MR patients to the Division of Human Genetics, UCT in order to stratify patient cohorts according to clinical presentation. With this in mind, together with the Divisional clinicians, a clinical manifestations datasheet is in the developmental stages. Furthermore it is envisaged that as the cost of array technology is reduced and more information on common XLMR gene mutations are collected that 'first-line' mutation detection will be performed on commercially available microarrays. Such technologies have been successfully employed in our Division for the group of retinal degenerative disorders (Roberts, Ramesar and Greenberg. 2009).

7.1.3 ACCOUNTING FOR UNEXPLAINED XLMR: FUTURE DIRECTIONS

Despite the nearly thirty years of research and the identification of 87 XLMR genes, the disease XLMR etiology remains unexplained in most families. The most comprehensively studied group of XLMR families are those of the EuroMRX consortium, comprising approximately 600 families. Mutation screening of 90 known or candidate XLMR genes in this cohort identified the molecular etiology in 42% of XLMR families (at least one obligate female carrier) and 17% of brother-pairs (2-5 sibs) (de Brouwer et al. 2007). More, recently a substantially more comprehensive study screened 718 Vega (Vertebrate Gene Annotation)-annotated X-chromosomal genes (at a coverage of 75%) in 208 XLMR family probands (Tarpey et al. 2009). A genetic cause of XLMR was attributed in a total of 25% of these XLMR families.

It is conceivable that additional mutations lie in the 25% of the Vega-annotated genes not screened in the large re-sequencing project (Tarpey et al. 2009). Also, additional gene mutations may lie in genic regions which have not yet been curated by the Vega project. Finally, given the number of NS-XLMR genes which have been implicated in single or minimal numbers of families, a substantial proportion of XLMR may be attributed to mutations in single genes making them difficult to detect. Future 'next-generation' DNA sequencing projects will most likely identify numerous monogenic causes of XLMR.

Despite the number of XLMR families hypothesised to be resolved by gene mutations in coding regions, it is also expected that a significant proportion of XLMR will be attributed to additional molecular mechanisms (Figure 7.1), including: Copy number Variations (CNVs) which have not yet been systematically investigated in XLMR cohorts, mutations in regulatory and noncoding RNA (ncRNA) regions, epimutations, or X-linked risk factors.

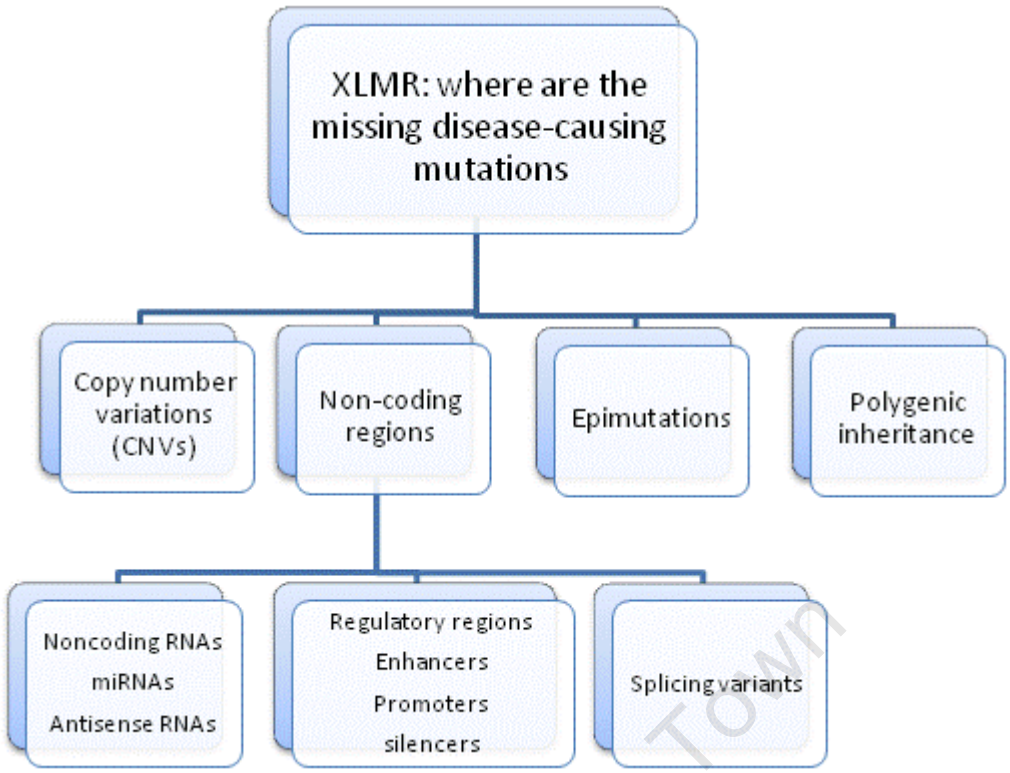


Figure 7.1 A hierarchal depiction of the putative regions or disease-mechanisms outside of the coding regions which may account for the pathogenesis of unexplained XLMR

7.1.3.1 Mutations in non-coding regions

Disease-causing variants in regulatory or mRNA splicing regions

The coding regions of X-chromosomal genes have been extensively investigated in XLMR patients in various studies. By comparison, relatively few have addressed the role of non-protein-coding X-chromosomal loci in the pathogenesis of XLMR. For instance functional chromosomal regions which encompass XLMR gene enhancers, promoters and consensus sequences critical for correct transcript splicing have not been strategically investigated.

The *ARX* gene may present a good candidate gene for investigation of regulatory region mutations. *ARX* is flanked by 11 UCEs which have been to hypothesised to function as enhancers (Bejerano et al. 2004), indeed to date one of these UCEs (uc.467) has been shown to function as an *ARX* enhancer (Colasante et al. 2008). In this research group at the Division four UCE elements downstream of *ARX* (including uc.467) were subjected to dHPLC for mutation detection in a small cohort of eight *FMR1*/*ARX* mutation negative patients, but no mutations were detected (M. Fish, submitted for BSc (Med) Hons dissertation). The cohort size was most-likely a limiting factor in this study;

however, now that techniques have been optimised for mutation detection, larger *FMR1*/*ARX* mutation negative cohorts can be screened for these mutations in these elements. It is also essential that future patient cohorts consist of XLMR patients with whom good communications exist, as, in the event of an UCE alteration being detected, a functional effect will need to be tested by assessment of *ARX* mRNA levels in the patient.

The encyclopaedia of DNA elements (ENCODE) project is a multicentre collaborative project which aims to identify all functional elements in the human genome (www.genome.gov/10005107) (ENCODE Project Consortium. 2004). The pilot phase of this project has been completed and future outcomes will provide valuable information for the sequencing of these regulatory variants (ENCODE Project Consortium et al. 2007). Furthermore, with the advent and application of technologies such as 'next generation' sequencing, especially being used in combination with microarrays for target sequence enrichment, the ability to perform these sorts of investigations are becoming possible.

Disease-causing variants in noncoding RNAs

Mutations of other non-coding regions may also affect gene regulation; these include the group of noncoding RNAs which include microRNAs (miRNAs), long noncoding (ncRNAs) (e.g. antisense transcripts) and a multitude of other small RNAs. To date, no ncRNA sequence mutations have been directly implicated in MR. However, dysregulation of miRNAs have been seen in a number of MR disorders including DiGeorge/22q11Del syndrome (Stark et al. 2008), Down syndrome (Kuhn et al. 2008, Kuhn et al. 2010), and Rett syndrome (Nomura et al. 2008), suggesting aberrations in this pathway may be important in MR pathogenesis. Only a single study by Chen and colleagues (2006) has attempted to test whether miRNA mutations are causative for XLMR (Chen et al. 2006). The group screened 13 brain-expressed miRNAs in 464 individuals with NS-XLMR and detected four sequence variants (in three miRNAs). These variants were determined to be functionally neutral and the authors conclude that mutations in these elements may be under strong selection pressure (Chen et al. 2006). However, as the location and sequence of novel miRNAs become known from the ENCODE data, additional studies in XLMR patient cohorts will elucidate whether the

findings by Chen et al (2006) can be validated and thus determine the pathogenic role of miRNAs in XLMR.

Mutations within long ncRNAs such as antisense transcripts have not been identified in MR patients. However, as is the case for miRNAs, there is empirical evidence to suggest these transcripts are dysregulated in patients with known MR syndromes. For example, the Antisense Fragile X Mental Retardation Protein (*ASFMR1* or *FMR4*) transcript is silenced in individuals with Fragile X syndrome (full mutation) and upregulated in premutation carriers (Ladd et al. 2007). However, siRNA-mediated silencing of *ASFMR1* did not affect *FMR1* expression, but rather, resulted in cell-cycle alterations and increased apoptosis (Khalil et al. 2008). It remains unclear though, whether this antisense transcript plays a role in the pathogenesis of FXS and requires further experimentation. A second antisense ncRNA transcript, the *UBE3* antisense transcript, has been shown to be upregulated in human and mouse MeCP2 deficient brains, with a concomitant reduction in *UBE3* expression. As UBE3 deficiency results in Angelman syndrome (AS) and given the phenotypic overlap between AS and Rett syndrome, it is proposed that dysregulation of the *UBE3* antisense transcript contributes to the clinical presentation of patients with MecP2 loss-of-function (Makedonski et al. 2005). While mutations in these ncRNAs have not been found to be causative for MR, their role in the pathogenesis of the disorder makes them good candidates for XLMR cohort screening, particularly in families where a linkage interval has already been defined. In addition, it is envisaged that projects such as ENCODE will determine the location and sequence of such antisense transcripts and enable researchers to identify putative disease-associations.

7.1.3.2 Polygenic models of inheritance

A hypothetical molecular mechanism which may account for the surplus of unidentified XLMR involves the adoption of a different mode of inheritance. It has been proposed that various gene polymorphisms on the X chromosome predispose males to MR, rather than cause the condition (Mandel and Chelly. 2004). This X-linked 'risk factors' hypothesis has recently been expanded to include autosomal polymorphisms (Raymond and Tarpey. 2006). Certainly the alterations identified in family Fx362 would seem to support this supposition; however, to date no significant empirical evidence has been forthcoming to prove this multifactoral model for the preponderance of male MR.

7.1.3.3 The role of epimutations

A significant proportion of XLMR may, in theory, be accounted for by the presence of epimutations in affected individuals. Epimutations are defined as epigenetic alterations that are not dependant on the primary DNA code, and that cause transcriptional silencing of a gene that is normally active, or conversely, activation of a silent gene (Holliday. 1987, Holliday. 2006). De novo epimutations have been reported in two patients with 14q32.2 hypomethylation who present with phenotypic presentation which mimics maternal uniparental disomy (UPD) of chromosome 14 (upd(14)mat) (Hosoki et al. 2008, Zechner et al. 2009). These reports suggest that DNA methylation errors may be a disease mechanism in sporadic cases of MR.

The biggest caveat to a trans-generational inheritance model for epimutations is that epigenetic marks are erased and reset in the germline at each successive generation. There are some challenges to this view. Evidence for the constitutional (germline) inheritance of epimutations has been demonstrated in plants (Chandler and Stam. 2004) and mice (Morgan et al. 1999). However, the only instance of constitutional epimutations in humans have been seen where the epigenetic alteration is dependent on the nucleotide sequence (Hitchins et al. 2007, Ligtenberg et al. 2009, Sutcliffe et al. 1992). Given that these alterations are essentially dependent on the primary DNA code, they do not strictly speaking constitute epimutations according to the accepted definition by Holliday (Holliday. 2006). However, there appears to be some ambiguity in current literature as the aforementioned reports refer to the changes in DNA methylation (that are based on the primary DNA code) as constitutional epimutations. , These alterations are referred to as epimutations herein, in keeping with current literature trends.

These constitutional epimutations include hypermethylation of the *MLH1* locus which occurs in the presence of a G>A SNP in hereditary non-polyopsis colorectal cancer (HNPCC) patients (Hitchins et al. 2007). Similarly *MSH2* promoter hypermethylation has been reported in HNPCC patients with *TACSTD1* 3' deletions (Ligtenberg et al. 2009). Finally, expansion of the *FMR1* CGG repeat causes promoter hypermethylation (Sutcliffe et al. 1992). However, to date, there is no evidence for a trans-generational inheritance of a 'true' epigenetic alteration.

The lack of identification of constitutional epimutations may also reflect the necessity for two 'hits' at biallelically expressed loci for disease manifestation. Indeed, those instances where epimutations have been associated with disease occur at imprinted loci (e.g. Angelman/ PraderWilli Syndrome) or in autosomal dominant disorders (e.g. HNPCC). However, given the hemizygosity of the X chromosome in males it may be expected that constitutional epimutations of X chromosomal gene regulatory regions may be more common. In order to test this hypothesis, clear, unexplained XLMR families should be investigated for DNA methylation aberrations across the X chromosome. Furthermore, these investigations could be guided by previously linked chromosomal regions in these families, e.g. of the EuroMRX consortium. It should be noted that while X-chromosomal changes to DNA methylation may be identified in such prospective studies, it is likely that these alterations will act in cis with an underlying change to the primary DNA code. However, these putative constitutive epimutations are probably more easily identifiable by DNA methylation alterations, with the putative genomic cause identified subsequently.

Finally, the *ATF7IP2* hypermethylation observed in the ATR-X patients described in this study, as well as previous studies showing DNA methylation alterations in these patients (Gibbons et al. 2000), add further credence to the role of epimutations such as DNA methylation in the pathogenesis of XLMR.

7.2 THE ROLE OF EPIGENETICS AS AN XLMR DISEASE MECHANISM

In this study, the hypothesis that DNA methylation alterations are a feature of MR in patients positive for XLMR gene mutations was tested. In order to fulfil this aim, DNA methylation profiles in an *ATRX* mutation positive family (XMR2) were investigated. This was achieved by conducting whole genome CpG island and promoter DNA methylation profiling using a MeDIP-Chip method. Subsequently, DMRs between affected and unaffected family members of family XMR2 were identified using a novel statistical pipeline. From the total of 113 putative DMRs identified, a single promoter region, encompassing *ATF7IP2*, was verified using bisulphite DNA sequencing. A semi-quantitative study at this promoter region using a combined T/A cloning and

bisulphite DNA sequencing technique, established that *ATF7IP2* promoter DNA methylation ratios in affected individuals vary between 8-34% as compared to unaffected sibs who have only 0-7% of sites methylated at the CpG sites within this locus. These results demonstrate that hypermethylation of *ATF7IP2* promoter region was apparent in *ATRX* mutation positive patients of this family.

The *ATF7IP2* hypermethylation observed in this study is hypothesised to lead to reduced levels of transcript and concomitantly, ATF7IP2 protein. The reduction in protein levels is likely to have an effect on the chromatin remodelling complexes in which ATF7IP2 functions. ATF7IP2 is known to bind MBD1 as well as the histone methyltransferase, SETDB1 and the transcription activator SP1 (Ichimura et al. 2005). This suggests that ATRX and ATF7IP2 may form part of the same transcription regulation complex and opens up new avenues of investigation for elucidating the pathogenesis induced by *ATRX* mutations. It can only be assumed that extended DNA methylation profiling in 'epigenetic' XLMR gene mutation positive patients will yield similar results. Elucidation of novel XLMR gene interacting proteins and subsequent investigation thereof will contribute to the understanding of XLMR disease mechanisms as well as indirectly enhance knowledge of the normal functioning of epigenetic mechanisms.

The altered DNA methylation profiles observed in *ATRX* mutation positive patients in this study therefore supports a role for dysregulation of this epigenetic mechanism in the pathogenesis of XLMR. The adoption of the whole-genome DNA methylation profiling methodology described in this dissertation to investigate the epigenetic profile of additional XLMR genes will elucidate the extent of this epigenetic disease mechanism. Those XLMR genes encoding proteins which interact with ATRX on a molecular level, for instance MeCP2 (Nan et al. 2007), would be good candidates for DNA methylation profiling in mutation positive patients. *MECP2* is also a particularly good candidate given its methylated DNA binding ability. However, evidence suggests that MeCP2 does not regulate target genes at promoter regions but rather through distal interactions (Yasui et al. 2007), therefore requiring a different array platform to that described in this dissertation.

While *ATRX* is expressed in the lymphocytes, the expression of a number of additional XLMR genes is confined to the brain. This is one of the major caveats to DNA

methylation studies such as those described in this dissertation as well as follow-up immunoprecipitation studies. The advent of induced pluripotent stem cell (iPS) technology (Dimos et al. 2008) will allow researchers to develop neuronal cell models which can circumvent this problem. Furthermore, generation of patient-specific neuronal cell lines, provide good cellular model systems for the testing of potential therapeutics.

An in depth knowledge of the disease mechanism governing the XLMR condition is the ultimate goal on the path to exploration of novel therapeutic targets and treatments for the disorder. For example, in the approximately thirty years since its first implication in XLMR, the *FMR1* gene has been extensively studied and the defective molecular mechanisms causing the FXS unravelled. These molecular mechanisms are discussed in detail elsewhere (D'Hulst and Kooy. 2009, Hagerman et al. 2009), but briefly summarised here. FMR1 is an RNA-binding protein which regulates long term depression (LTD) of synapses through the glutamate receptor (mGluR) pathway. In the absence of FMR1, this mGluR pathway lacks the negative feedback loop induced by FMR1 and is hence over-stimulated. The resulting exasperated neuronal LTD eventually leads to weakened synaptic connection in FXS patients. Intricate knowledge of functioning of this system can be exploited in the quest to identify therapeutic targets.

7.3 TREATMENT OF XLMR: THE ULTIMATE GOAL

The development of the aforementioned mGluR theory has led to the suggestion that antagonists of this pathway may be utilised in the treatment of FXS. Indeed, already significant phenotypic rescue has been seen in mouse, fly and zerbrafish models of FXS (McBride et al. 2005, Tucker, Richards and Lardelli. 2006, Yan et al. 2005). Given the successes of this pharmacological intervention in animal models, in 2008 the first clinical trial in FXS patients using an mGluR antagnostic (Fenobam) was established (Hagerman et al. 2009). Also, in a recent study a small group of FXS patients treated with Lithium showed improvement in verbal memory and behaviour, possibly due to a correction of the mGluR pathway functioning (Berry-Kravis et al. 2008). Finally, other studies have suggested that GABAergic antagonists may also be effective treatment in FXS (Chang et al. 2008). Clinical trials testing amelioration of some of the symptoms of FXS using this antagonist are scheduled (Cornish, Turk and Hagerman. 2008).

The aforementioned studies suggest the therapeutic correction of these neurotransmission signalling pathways is a promising treatment protocol for FXS patients. However, given the dynamicity of epigenetic mechanisms, therapeutic interventions which could target erroneous or pathogenic marks of this mechanism could provide novel targets for treatments. For instance, an individual with a full *FMR1* expansion mutation was recently identified with normal intelligence (Pietrobono et al. 2005). It was shown that the absence of both *FMR1* hypermethylation and H3K4 methylation resulted in the production of normal transcript levels. However, reduced translation was observed due to the CGG expansion with protein levels reduced in the individual by 30% as compared to normal controls. Concomitantly, a separate study demonstrated that 5-aza-deoxycytidine treatment of FXS full mutation cell lines reversed *FMR1* hypermethylation as well as the associated repressive histone marks (Pietrobono et al. 2002, Tabolacci et al. 2005). Collectively these results suggest that treatment with similar DNA demethylation targeting drugs may also be useful therapeutic treatment in FXS patients, though of course off target effects would need to be assessed.

MECP2 is a well-studied XLMR gene, second only to *FMR1*; however, to date no therapies have been established for either loss-of-function mutations in Rett Syndrome patients, or overexpression mutations in XLMR patients. MeCP2 function is thought to be particularly complex with this protein being involved in both activating and repressive chromatin complexes (Chahrour et al. 2008, Nan et al. 1998, Yasui et al. 2007). Given that the Rett Syndrome phenotype was partially rescued after onset of the disease in mouse models suggests that this disorder may be amenable to therapy (Giacometti et al. 2007, Guy et al. 2007, Jugloff et al. 2008). To this end, Rett syndrome gene therapy is being developed using *MECP2* specific self-inactivating retroviral vectors and has shown success in neural stem cells (NSC) (Rastegar et al. 2009). However, given MECP2's reported epigenetic function as well as the dynamicity of epigenetic mechanisms, therapies which target this pathway may be explored. Possible avenues of investigation may include drugs such as Trichostatin A (TCA) which inhibit HDACs (class I and II) (known to interact with MeCP2). A recent report showed that cocaine self-administration in mice resulted in the up-regulation of several genes including *MECP2*, *HDAC2* and *HDAC11*, this up-regulation was corrected by treatment with TCA (Host et al. 2009). These results suggest that TCA and additional available HDAC inhibitors could be explored as a novel therapeutic

approach in patients with *MECP2* overexpression mutations (male *MECP2* duplication carriers).

The molecular mechanisms underlying the two aforementioned XLMR genes, *FMR1* and *MECP2* are perhaps the most thoroughly investigated. Hence, particularly in the case of *FMR1* potential therapies are being actively pursued. Concomitantly, a thorough understanding of the molecular mechanisms of other XLMR genes will pivotal in the pursuit of potential targeted therapies. This is especially true for the 'epigenetic' XLMR genes such as *ATRX*, patients with mutations in these chromatin remodelers may be good candidates for therapy given the dynamic nature of epigenetic mechanisms.

7.4 CONCLUDING SUMMARY

The principal findings of this dissertation were:

- The identification of three disease-causing CNVs: Xq26.3-27.3Del in family Fx444, Xq25Dup in XMR8 and 17p11.2 in Fx343 (Chapter 2)
- Definition of a disease-associated interval in family Fx67, which putatively harbours a mutation in a novel XLMR gene (Chapter 3, 4)
- Definition of a linked critical interval in family XMR2 and subsequent detection of a disease-causing mutation (c.5987_6011del) in positional candidate gene, *ATRX* (Chapter 3, 4)
- The identification of two *ARX* disease-causing mutations (c.428_451dup) in families Fx391 and Fx444 (Chapter 5)
- The identification of a putative *KDM5C* mutation (c.2517-7_8insTAC) in individual Fx361.1 (Chapter 5)
- The identification of hypermethylation at the *ATF7IP2* promoter region in *ATRX* mutation positive family, XMR2 (Chapter 6).

The principal findings of this study contribute to numerous facets of the XLMR/MR research field. Firstly, the identification of six (putatively seven) disease-causing mutations allows further delineation of the molecular causes of MR/XLMR. Secondly, the elucidation of a genetic cause of MR in these families enhances clinical management of this disorder in these patients as well as genetic management of the families. Finally, the role of DNA methylation as a disease-mechanism has been

substantiated and thus provides the foundation for future research directions concerned with understanding disease pathogenesis.

The identification of a low penetrant Xq26.3-27.3Del mutation in family Fx444 highlights the importance of developing guidelines for the genetic counselling and molecular diagnosis of patients carrying similar variably penetrant CNVs. The Xq25Dup identified in XMR8 has led to the detection of novel XLMR candidate genes (*THOC2* and *STAG2*). These results illustrate the importance of CNV analysis in both the molecular diagnosis of MR as well as novel XLMR/MR gene detection.

The identification of a disease-associated interval in family Fx67 and the absence of mutations in the two known XLMR genes suggested that this family harbours a mutation in a novel XLMR gene. These results illustrate the importance of molecular investigations in new XLMR families and exemplify the contribution to the delineation of unexplained MR that research in developing countries will make in the future.

The elucidation of a disease-causing mutation in *ATRX*, as well as two *ARX* mutations now affords enhanced clinical management of these patients as well as genetic counselling, carrier testing and prenatal diagnosis for these families. Also, there is now compelling evidence that the *KDM5C* mutation (c.2517-7_8insTAC) described in this study is a disease-causing mutation.

As a result of these investigations, two diagnostic tests (for the two common *ARX* mutations and the *ATRX* mutation hotspots) have been incorporated into the diagnostic protocols offered by the NHLS, based at UCT Medical School. In addition, the logistics and feasibility of including CNV detection using microarray analysis into the diagnostic procedures of the NHLS is in the planning stages. The expansion of the repertoire of diagnostic tests offered to MR patients will aid clinicians in South Africa to attribute a molecular diagnosis to an increased number of MR patients. In turn, a molecular diagnosis will facilitate clinical management of the disorder, as well as genetic counselling, carrier testing and prenatal diagnosis in these families.

While the DNA methylation alterations in family XMR2 were not widespread, the detection of hypermethylation at *ATF7IP2* suggest that these aberrations are a molecular feature of MR for at least one XLMR gene, *ATRX*. Furthermore, ATF7IP2 was

identified as a novel potential interacting protein of ATRX. Given the role for ATF7IP2 in chromatin remodelling further investigation of the relationship between ATRX and ATF7IP2 could provide important clues to understanding the pathogenesis of ATR-X syndrome. This dissertation has therefore laid the foundation for future research directions in understanding *ATRX* mutation pathogenesis. Furthermore, these investigations suggest *ATF7IP2* may be a good candidate gene for mutation screening in patients with chromosomal rearrangements in this region, as well as potential future autosomal recessive families linked to the region by homozygosity mapping. Finally, hypermethylation at the *ATF7IP2* promoter region may hold promise as a future molecular diagnostic test for the presence of an *ATRX* mutation.

In conclusion, the DNA methylation methodology described in this study can be applied to additional lymphocyte-expressed 'epigenetic' XLMR genes. These future investigations will elucidate the extent of the DNA methylation alterations in MR patients, identify novel interacting proteins or protein targets and finally, lead to the discovery of novel MR candidate genes. The acquisition of intricate knowledge of epigenetic defects in MR patients, will ultimately contribute to understanding the role these epigenetic mechanisms play in learning and memory, the cornerstones of cognitive functioning.

APPENDICES

APPENDIX 2A: DNA CONSENT FORM

REQUEST FOR MOLECULAR STUDIES (DNA)

Molecular Laboratory

Division of Human Genetics
IIDMM, LEVEL 3
UCT Medical School, Observatory 7925

Tel: (021) 406 6425 Fax: (021) 406 6826

Blood should be drawn in 2 plastic EDTA Tubes (Purple top) +/- 10ml each using a yellow barrel. Each tube should be inverted to mix and should be clearly labelled with the patient's name and DOB
Keep blood in fridge at 4°C until able to send to laboratory

Please DO NOT send specimens on ice or frozen.

Please fill in all the information requested:

Surname:_____First Name(s):_____

New Family: Yes ☐ No ☐ (If no, please fill in family name) Family name: _____

Medical Aid: _____Medical Aid No: _____

Sex: M ☐ F ☐ Date of Birth: Year:_____ Month:_____ Day: _____

Number of children: _____

Ethnic Origin : (please indicate ancestry of both your mother and father)_____

Contact Address: _____ Town: _____ Fax: Tel: _____

Referring Doctor/Sister: _____ Town: _____ Fax: Tel: _____

Hospital or Address: _____ Town: _____ Fax: Tel: _____

Reason for Referral (Clinical diagnosis):

Affected ☐ At Risk ☐ Carrier ☐ Spouse ☐ Query ☐ Unaffected ☐

Becker Muscular Dys. ☐	Duchenne Muscular Dys ☐	Colonic Carcinoma ☐
Fragile-X Syndrome ☐	Bipolar Disorder ☐	Huntington Disease ☐
Retinitis Pigmentosa ☐	Spinocerebellar Ataxia ☐	Waardenberg Syndrome ☐

Additional disorders (apparent or previously treated): _____

Additional family history _____

Clinical Details:

Physical disability ☐ Mental retardation ☐ Deafness ☐ Impaired vision ☐ Night blindness ☐

Other: _____

Have samples from this patient been sent to a DNA lab before? (DELETE WHERE NOT APPLICABLE) YES / NO / Don't Know

If Yes, where: _____

For Laboratory use only:

DNA number: _____ *Vol.Blood:* _____*(ml)* *Other:* _____

Date Received: Year: _____ *Month:* _____ *Day:* _____ *Computer Index No:*_____

CONSENT FOR DNA ANALYSIS AND STORAGE

1. I, _____ , request that an attempt be made using genetic material to assess the probability that: I / my child / my unborn child (DELETE WHERE NOT APPLICABLE) might have inherited a disease-causing mutation in the gene for:_____

2. I understand that the genetic material for analysis is to be obtained from: blood cells/skin sample/other (specify) (DELETE WHERE NOT APPLICABLE) :

3. I request that **no** portion of the sample be stored for later use. ☐ **(**MARK IF APPLICABLE)

   Or

   I request that a portion of the sample be stored indefinitely for (DELETE WHERE NOT APPLICABLE):

   (a) possible re-analysis
   (b) analysis for the benefit of members of my immediate family
   (c) research purposes, subject to the approval of the University of Cape Town Research Ethics Committee, provided that any information from such research will remain confidential.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doctor, in accordance with the relevant protocol, if and when available.
   In addition, I authorise that they may be made known to: (DELETE WHERE NOT APPLICABLE) :
   other doctors involved in my care
   the following family members: _____
   other: _____

5. I authorise / do not authorise my doctor(s) (DELETE WHERE NOT APPLICABLE) to provide relevant clinical details to the Division of Human Genetics, UCT.

6. I have been informed that:
   (a) there are risks and benefits associated with genetic analysis and storage of biological material and these have been explained to me.
   (b) the analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.
   (c) the genetics laboratory is under an obligation to respect medical confidentiality .
   (d) genetic analysis may not be informative for some families or family members.
   (e) even under the best conditions, current technology of this type is not perfect and could lead to incorrect results.
   (f) where biological material is used for research purposes, there may be no direct benefit to me.

7. I understand that I may withdraw my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERED BY:**

_____ DATE: _____

Patient signature _____ **Witnessed consent** _____

NOTE - PLEASE INSERT A FAMILY PEDIGREE DRAWING ON THE REVERSE OF THIS FORM

APPENDIX 2B: CLINICAL FEAUTURES OF PATIENTS SELECTED FOR WHOLE-GENOME CNV ANALYSIS

	FAMILY	INHERITANCE	EHTNICITY	PHENOTYPE
1.	FX76 And XMR3	XLMR	M	**AFFECTED MALE 1(FX76.2):** Severe MR, Global developmental delay, Delayed speech, Asymmetrical hearing loss, Epileptic seizures, Ataxia, Small stature, Wheelchair bound Gross gastric reflux and delayed gastric emptying, Chronic URTI Splenomegaly (may relate to respiratory infection), hepatomegaly, Ear infections – grommets Behavior: Poor eye contact, Aggressive, Limited social interaction, Autism spectrum **AFFECTED MALE 2 (XMR3):** Severe MR, Unable to speak, Coarse facial features **MOTOR:** Epilepsy, Jerking eye movements, Hypotonic bilaterally, Hyporeflexia, "intractable" seizures, "floppy" no spontaneous movement, Contractures, Reflexes appear to be decreased on left side, 5yrs relatively fit-free period during which he was on a small dose of Valproate, Twitching, Todd's paralysis in left side, Pyrexial fits **INFECTIONS:** Pneumonia, Hepatitis, severe prolonged encephalopathic illness, Prolonged chest infections Hyperactive Metabolic encephalopathy EEG left temporal abnormality
2.	FX131	XLMR	M	Severe MR Mild hearing loss Large depigmented area on abdomen No eye contact Autism Possibility of hypopituatarism Bone of wrist in left hand appear amorphic Very poor speech development
3.	FX141	XLMR	M	Speech delay Hearing problems Behavioural problems
4.	FX143	XLMR	M	Severe MR Speech delay Behavioural problems Dysmorphic features: Big ears, Simple ears "shawl scrotum" Epilepsy – drop attacks Anaemic

	FAMILY	INHERITANCE	EHTNICITY	PHENOTYPE
5.	FX151	XLMR SIB PAIRS	M	Moderate MR Speech delay Microcephaly Short stature Brisk reflexes Behavioural problems
6.	FX181	XLMR	M	Moderate mental retardation Speech delay Behavioural problems Short stature
7.	FX233	XLMR	M	Behavioural problems but not hyperactive Epilepsy Microcephaly Brisk reflexes Dysmorphic features – large ears, Low set ears, Epicanthic folds
8.	FX256	XLMR	C	Premature at 36 weeks Suspected patent ductus arteriosis after birth Tall stature Single seizure episode Dypsmorphic features -prominent forehead, Prominent metopic ridge, Up-slanting palpebral fissures, Down turned corners of mouth, Small nose with anteverted nares, Long philtrum, Heterochromia of iris, Low set ears, Simple ears Undescended testes
9.	FX295	XLMR	M	Dysmorphic features - epicanthic folds, Long face, Arched brows, Prominent lips, Prominent ears Large hands Macro-orchidism
10.	FX317	XLMR		Severe mental retardation Speech delay Behavioural problems Autistic features

	FAMILY	INHERITANCE	EHTNICITY	PHENOTYPE
11.	FX338	XLMR		Speech delay Microcephaly Dysmorphic features – simian crease, Bat ears, squint Café au lait spot No behavioural problems
12.	FX343	Isolated male		small penis behaviour hyperplegia high pain tolerance
13.	FX361	XLMR SIB PAIRS	M	History of diarrhoea and vomiting Myoclonic jerks Speech delay - expressive Cerebral interdigitations present in anterior interhemisphere area Seizures Dysmorphic features: Epicanthic folds, Palatine malformation: deep with a ridge in the middle, Hypotelorism
14.	FX382	XLMR		Severe mental retardation Speech delay Hyporeflexia Hypotonia High arched palate Dysmorphic features – frontal bossing, Coarse facies, Malar hypoplasia, Bulbous nose, Micrognathia. Ptosis, Microcephalic, Squint Marfenoid habitus – tall, Long fingers, Cubitus valgus
15.	FX398	XLMR	M	Speech delay Epilespy Hyperactive Behavioural problems Hypothyroidism – subclinical Dysmorphic – hypertelorism, Micrognathia, Epicanthic folds Hyperreflexia

	FAMILY	INHERITANCE	EHTNICITY	PHENOTYPE
16.	FX403	XLMR	C	Delayed walking Speech delay Recurrent ear infections and URTI Mild obesity Generalized hypotone Developmental coordination disorder
17.	FX444	Isolated male		profound mental retardation Growth retardation: height, weight head circumference all in 3rd centile Seizures (Lennox Gestualt epilepsy syndrome) Short broad feet and hands with clinodactyly Undescended testes hypotonia Haemophilia B No speech or ambulation
18.	FX497	Isolated male	M	Dyplexia Machroorchidism
19.	FX509	Isolated male	M	dysmorphism – fairly prominent ears and prominent incisor teeth absence of speech Unsteady gait and increased limb tone tonic-clonic seizures
20.	FX512	XLMR		Dysmorphic features - tented upper lip, large tongue, Prominent lower lip, Protruding ears, Hypotonia
21.	FX516	XLMR SIB PAIRS	C	Facial dysmorphism CT: Dandy walker cysts? Posterior fossa malformation: megacisterna magna or posterior fossa arachnoid cyst
22.	FX517	XLMR		Speech delay Hyperactive Hypotonia Dysmorphic – large ears, Prominent jaw, Anti-mongoloid eyes, Deep-set eyes, Broad nasal bridge, Short neck, squint, "widows peak" Mild brachydactyly Height, weight and head circumference all normal Delayed bone age
23.	FX520	Isolated male	M	behavioural problems: ADHD, aggressive dysmorphic features ?Williams syndrome

	FAMILY	INHERITANCE	EHTNICITY	PHENOTYPE
24.	FX523	XLMR	M	Seizures Clinodactyly Subtle dysmorphism Microcephaly Small stature
25.	XMR8	isolated	M	Severe mental handicap **SKULL:** microcephaly, Abnormal skull bones, Metopic suture open, 3rd fontanelle, Trigonocephaly **BRAIN:** Large ventricles anterior, posterior positive, bilateral ependymal cysts, **FACIAL DYSMORPHISM:** coup shaped mouth, flattened nasal bridge, hockey stick palmar crease, 3 hair 'crowns', Anterior sagittal cleft deep to root of nose, Narrow forehead, Narrow palate, Micrognathia, Flattened malar area, squint **GENITOURINARY**: Undescended testes, Both testes intra-abdominal, microgenitalia, disconnected vas, Umbilical hernia **MOTOR:** GOR, Vomiting, problems swallowing, Delayed brisk reflexes, Increased tone in arms, Hyperreflexia, Slight muscle hypotone, Not walking, Tremendous sensory aversion of feet, Reciprocal crawling, Triplegic, no definite spasticity **SKELETAL**: Right arms tends to be retracted at shoulder **BEHAVIOUR**: Aggressive and hysterical, irritable, frustrated. **OTHER:** Chest infections, Heart murmurs, Speech delay (no speech), Seizures
26.	XMR 12	Isolated male	B Xhosa	Seizures Microcephalic Dysmorphism: Broad forehead, brachycephaly, impression of mild hypertelorism, broad nasal root and tip CT brain: "classic lissencephaly". Thick agyric cortical mantle posteriorly. Low white matter but normal myelination.
27.	XMR 13	Isolated female	B Xhosa	Dysmorphic facies reminiscent of Noonan Syndrome ventricular septal defect Recurrent hypoglycaemia Septicaemia
28.	XMR 14	isolated female	M	Doliocephaly Pulmonary Stenosis Dysmorphic Hearing problems Seizures
29.	ANG 13	Isolated male	C	Angelman syndrome? dysmorphic

*M= Mixed ancestry, C=Caucasian, B=Black

APPENDIX 2C: ALL PRIMER PAIRS USED TO VALIDATE PUTATIVE CNVS

Sample	location	primer name	sequence	Sample	location	primer name	sequence
Fx151	6q24.2-3 dup	UTRN_E54_F	agggagctatggatgacctg	Fx277	Xp11.22-21 dup	JARID1C_EX1_F	tatacgaggctcggaaggac
		UTRN_E54_R	tgcagcgagtcaatgagtaag			JARID1C_EX1_R	tcttggtttgtcagcgtctc
		EPM2A_E2_F	aaacaacttggtggatggtg			WNK3_E2_gF	tctcctagtggcagattcctg
		EPM2A_E2_R	gcttcatttcattggtgtgc			WNK3_E2_gR	cctcaacccatgtttcagtg
		SHPRH1_E9_F	cacctctgattaccgctttg			FGD1_5'upstream_F	gaatagttttccccagcaatg
		SHPRH1_E9_R	gttgccacaggtgacacttc			FGD1_5'upstream_R	accctcagttccctgatttg
		GRM1_E2_F	tctgtatcgcccattctgac			GNL3L_3'UTR F	gtattcaggggcaaccaaag
		GRM1_E2_R	gaagcctctctcggagtttg			GNL3L_3'UTR R	agggttaacggtggtaatgg
		STXBP5_E23_F	ctccgctcttcatttctgtg			USP51_g3'downstream_F	catgcctagttccctgaacc
		STXBP5_E23_R	gcaatgacaagcactgttcc			USP51_g3downstream_R	ggaggagtaaaagccaccag
Fx233	14q32.12 dup	BTBD7_E9_F	tgcttgcagaccttctactcc	Fx343	17p11.2 del	NCOR1_gE30_F	caccaccatcaaagaaatgg
		BTBD7_E9_R	ttggtgatgccttattgtgg			NCOR1_gE30_R	ctggagttttccgactttcc
		ATXN3_E8_F	tggttgcagttattaccagtgc			FLCN_gE14_F	cgacccaaagaggacacac
		ATXN3_E8_R	gcaaatcctcctcatcttcg			FLCN_gE14_R	tccagaacttcagcagcttg
		NDUFB1_E1_F	gtatgatttgctggcgtcac			SMCR7_gE1_F	gctccgattggagttaggg
		NDUFB1_E1_R	cgagaccaagggcaacag			SMCR7_gE1_R	ctgggaatggaacgtgaag
		CPSF2_E15_F	cttcaagcaagttctcttacgg			SLC5A10_gE7_F	gaccagatcggtggttacg
		CPSF2_E15_R	cggactgctacttgattgttg			SLC5A10_gE7_F	caggtggtgttggcaatg
		RPS6KA5_E16_F	ggcttgaggtacaatgaatgg			MAP2K3_gE10_F	atcctgcggttcccttac
		RPS6KA5_E16_R	tgctttcacacaggtatgcac			MAP2K3_gE10_R	caaactcgggggagaaac

Sample	location	primer name	sequence	Sample	location	primer name	sequence
Fx444	Xq26.3-27.3 del	ARHGEF6_E22_F	gaagcaatgcctggaagaag	Fx512	Xp11.3 dup	CASK_E15_F	attgagggtcacacctcctc
		ARHGEF6_E22 _R	actggactcgcctcgaatac			CASK_E15_R	tgaaactgtaccagccgaac
		FGF13_E6_ F	tccatgagccacaatgaatc			EFHC2_5'UTR_F	actgttggcgacaagaattg
		FGF13_E6 _R	acttttgggtgaaggactgc			EFHC2_5'UTR_R	tgcacacacaaagattgactg
		F9_5'UTR_F	cgaccttaccactttcacaatc			FUNDC1_3'downstream F	cagggctggtatcatacttcc
		F9_5'UTR_R	tgatgaggcctggtgattc			FUNDC1_3'downstream R	gcttttcccaattccaacc
		SOX3_E1_F	gcaatgtacagccttctggag			UTX_3'UTR_F	gcaccactggtttttgtagc
		SOX3_E1_R	ctgcgttcgcactactcttg			UTX_3'UTR_R	cgtccagttggttgacactc
		LDOC1_E1_F	gaaccgattctgcaacgac			SLC9A7_E17_F	tgtttcccctggaagataatg
		LDOC1_E1_R	ggatggggctatccatctc			SLC9A7_E17_R	aacaacactagggcttgcag
		SPANXN4_E2_F	cgtttacctgctgctcctg	Fx512	Xp22.11 Dup	MBTPS2_E7 _F	ccttttgtgggagaccttg
		SPANXN4_E2_R	tgctgtgcatctacccagtc			MBTPS2_E7 _R	aaccaatttggggctcatag
		UBE2NL_3'end_F	cagtcagcataggcaaagagtc			ACOT9_E17_ F	acccacagtgacgtggtatc
		UBE2NL_3'end_R	catcttttgacgccagtcc			ACOT9_E17_ R	gctgaatacatcctcctcctg
		SLITRK2_E2_F	tttgccccttcctatgaatc			KLHL15_E3_ F	gtggtgaagttctgctgctc
		SLITRK2_E2_R	gtggttgggttttgactgc			KLHL15_E3_ R	tgactccctcgatgtttacg
		Xq27.3_1_ F	acctgactttgatctgtcactacc			POLA_E9_ F	gcaatggagtttgaagatgg
		Xq27.3_1_ R	ccctccttggctcttacttg			POLA_E9_ R	ctgctggctcactctctttg
		Xq27.3_2_ F	catttgccaacggagtaatg			MAGEB6_5 upstream _F	aggaccagccaggtttagag
		Xq27.3_2_R	tgctcaaaggtgaagtttgg			MAGEB6_5 upstreamR	ctcccaatcttttcgtgttg
		FMR1_5'end_F	tgccaagcagtgtaatgagg				
		FMR1_5'end_R	ctgtggtgaaaacccttgtg				

Sample	location	primer name	sequence	Sample	location	primer name	sequence
FX517 XMR14	Xq25Del	ZDHHC9_E11_F	ccttttaggaatgggacagg	XMR8	Xq28Dup	THOC2_E23_F	tgtttccaaagtctgggatg
		ZDHHC9_E11_gR	tgtcagagtggatgggagac			THOC2_Ex23_R	tcgttcatagctggtgtgtg
		AIFM1_5'upstream_F	aggtcagggattcaaacagc			STAG_Ex20_F	ggatgccttattgcgacag
		AIFM1_5'upstream_R	ctcagcagaaatggatttgg			STAG_Ex20_R	actcttcattacagagtgcatgg
		SUHW3_3' upstream_F	aaagaacagggaaatctgcac			STAG2_Ex34_gF	ccccctctctctctcattagg
		SUHW3_3' upstream_R	tgcacattcgtcacaacatc			STAG2_Ex34_gR	ccacttagaaaatgacttcaccac
		SUHW3_5'UTR_F	caccaaatagagcacaacttcc			ODZ1_Ex3_F	ggaagaaaaccaagacagtcatac
		SUHW3_5'UTR_R	tgaagcactgatctcatggac			ODZ1_Ex3_R	ggctattgtaattcatcctcagc
		GPR119_3''downstream_F	aatgacaaggttggcagatg	XMR12	4p16.2Del	RGS12_Ex2 _F	gggctacttaggctccattg
		GPR119_3'downstream_R	gaaaggcagagagtggttcc			RGS12_Ex2 _R	cgagtggattttctgctctg
XMR8	Xq28Dup	CUL4B_gE3_F	cctttacaacccagggattc			BC042381_Ex1 _F	gaacctggagagagccagag
		CUL4B_gE3_R	acgcagcttcttctgtatcg			BC042381_Ex1 _R	gttgacctttcccaacgtc
		GRIA3_gE11_F	cccctggcttatgaaatctg			ADRA2C_Ex3 _F	ttagagagcagtggcagagg
		GRIA3_gE11_R	ttgtcttccaagtgccattc			ADRA2C_ Ex3_R	ccgactaggtcctggagaag
		GRIA3_Ex12_F	aatctcgctgctttcctgac			TMEM128_Int2 _F	gcctagaatgagctgttggag
		GRIA3_Ex12_R	ttcttttgttgaaccggagtc			TMEM128_Int2_R	tggaggctgagatgtgtttg
		GRIA3_Ex13_F	ttacatgaaatcagcggagc			TMEM128_Int2_R	Tggaggctgagatgtgtttg
		GRIA3_Ex13_R	cattcatggttgactccagc				
		GRIA3_Ex14_F	ctccgggagtaaggtcagtc				
		GRIA3_Ex14_R	gtgtgtgttgggaaggaaag				
		GRIA3_Ex15_F	ctctgagcctgagcaatgtg				
		GRIA3_Ex15_R	Ctgcccgtgatttgtaacag				

APPENDIX 2D: PRIMER PAIRS USED FOR MICROSATELLITE MARKER ANALYSIS IN FAMILY FX444

Table 2C Details of the X-chromosomal microsatellite markers used for haplotype analysis in family Fx444

Marker	Location on X-chromosome (Mb)	Heterozygosity	Size (bp)	Fluorescent tag (i.e. M13 forward primer used)
DXS1071	1.7	-	110-130	Fam
DXS1223	8.3	0.75	150-160	Fam
DXS8022	13.7	0.81	190-210	Fam
DXS1202	26.2	0.81	240-260	Hex
DXS8090	36.7	0.81	260-270	Ned
DXS1003	46.3	0.80	160-180	Ned
DXS1199	53.6	0.72	340-360	Hex
DXS8092	73.9	0.87	140-160	Rox
DXS990	92.8	0.77	160-180	Hex
DXS1220	114.4	0.71	190-210	Fam
DXS1212	122.2	0.73	320-340	Rox
DXS1192	138.1	0.84	190-210	Rox
DXS1193	148.1	0.75	170-190	Ned

APPENDIX 2E: ALL CNVS DETECTED IN 30 MR PATIENTS USING THE 250K NSP AFFYMETRIX ARRAY

	PATIENT	DUP/ DEL	chromosome	BAND	SIZE (Mb)	START		END		ACTION
						SNP id	Chromosomal position	SNP Id	Chromosomal position	
1.	FX76	DUP	14	q11.2	0.16	rs4412905	19,336,854	rs10141075	19,489,991	CNP
2.	FX76	DUP	15	q11.2	0.98	rs4931862	18,427,103	rs1346662	19,407,629	CNP
3.	FX76	DEL	18	q12.1	0.007	rs6507033	29 277 562	rs6507040	29 285 278	No genes in region
4.	FX131	DUP	7	q11.22	0.58	rs517781	70,638,772	rs17147106	75,859,448	Not CNP, XLMR family
5.	FX131	DEL	8	q24.23	0.17	rs894089	137,861,009	rs7842167	137,931,617	CNP
6.	FX131	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
7.	FX131	DUP	15	q13.3	0.42	rs7174213	29,806,023	rs4238560	30,231,488	CNP
8.	FX141	DUP	4	q28.2	0.7	rs2391537	131 081 327	rs9917861	131 775 687	CNP
9.	FX141	DEL	5	p13.1	0.06	rs1363883	41,293,670	rs399760	41,300,320	CNP
10.	FX141	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
11.	FX141	DUP	17	q21.31	0.2	rs17659881	41,518,102	rs2732615	41,719,833	CNP
12.	FX143	DUP	11	q23.3	0.94	rs1624049	115 542 441	rs12271161	116 485 121	Not CNP, XLMR family
13.	FX151	DEL	2	q37.3	0.08	rs7600642	242,636,531	rs6759916	242,717,659	CNP
14.	FX151	DUP	6	q24.3	1.07	rs581234	145 649 820	rs9403788	146 827 677	Verify by gQPCR
15.	FX151	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
16.	FX181	DEL	1	q31.3	0.15	rs424535	193,462,805	rs4915559	193,614,160	CNP
17.	FX181	DUP	11	p15.1	0.34	rs1384649	18,896,898	rs7104587	18,930,870	CNP
18.	FX181	DUP	15	q11.2	0.68	rs4931862	18427103	rs8040821	19112164	CNP
19.	FX233	DUP	14	q32.12	0.3	rs4900092	91 499 257	rs2402130	91 870 956	Verify by gQPCR
20.	FX233	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
21.	FX256	DUP	1	q44	0.53	rs11204636	244,818,450	rs2511	245,353,397	Not CNP, XLMR family

	PATIENT	DUP/ DEL	chromosome	BAND	SIZE (Mb)	START		END		ACTION
						SNP id	Chromosomal position	SNP Id	Chromosomal position	
22.	FX256	DUP	11	p15.1	0.34	rs1384649	18,896,898	rs7104587	18,930,870	Not CNP, XLMR family
23.	FX277	DEL	5	q23.3	0.01	rs7719223	127,852,880	rs6595831	127,863,541	Not CNP, XLMR family
24.	FX277	DUP	6	q21	0.53	rs17065302	105,336,339	rs1211490	105,864,431	CNP
25.	FX277	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
26.	FX277	DUP	17	q21.31	0.2	rs17659881	41,518,102	rs2732615	41,719,833	CNP
27.	FX277	DUP	X	p11.22	0.7	rs2495782	54,081,339	rs5960365	54,787,066	Verify by gQPCR
28.	FX295	DUP	2	q14.3	0.11	rs4662794	128,748,055	rs4662808	128,861,934	Not CNP, XLMR family
29.	FX295	DEL	10	q11.22	0.05	rs10906958	47030119	rs7088096	47079558	CNP
30.	FX295	DEL	12	p12.1	0.05	rs16926687	23778139	rs11047098	23835465	Not CNP, XLMR family
31.	FX295	DUP	17	q12	0.5	rs1614133	31,429,427	rs2277662	31,923,810	CNP
32.	FX317	DEL	9	p24.1	0.09	rs1658957	6,662,097	rs7861740	6,759,429	CNP
33.	FX317	DUP	11	p15.1	0.03	rs1384649	18,896,898	rs7104587	18,930,870	CNP
34.	FX317	DUP	14	q11.2	0.16	rs4412905	19,336,854	rs10141075	19,489,991	CNP
35.	FX317	DUP	22	q11.21	0.4	rs2252257	17,014,854	rs10854539	17,412,288	CNP
36.	FX338	DUP	9	q33.3	0.11	rs3850586	126,566,536	rs6478760	126,682,300	Not CNP, XLMR family
37.	FX338	DUP	14	q31.3	0.1	rs17730574	83,607,255	rs7151936	83,704,489	Not CNP, XLMR family
38.	FX343	DEL	17	p11.2	3.2	rs9910842	16,486,774	rs188138	19,689,696	Verify by gQPCR
39.	FX361	DEL	11	q11	0.14	rs2868510	55,217,364	rs7114700	55,364,837	CNP
40.	FX361	DUP	14	q11.2	0.16	rs4412905	19,336,854	rs10141075	19,489,991	CNP
41.	FX361	DUP	15	q11.2	0.68	rs4931862	18427103	rs8040821	19112164	CNP
42.	FX361	DUP	17	q21.31	0.12	rs17659881	41 521 621	rs2732615	41 719 833	CNP
43.	FX382	DUP	8	q11.22	0.54	rs7012121	48,124,531	rs6988862	48,669,181	CNP
44.	FX398	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
45.	FX398	DUP	17	q21.31	0.12	rs17659881	41 521 621	rs6640123	41 719 833	CNP

	PATIENT	DUP/ DEL	chromosome	BAND	SIZE (Mb)	START		END		ACTION
						SNP id	Chromosomal position	SNP Id	Chromosomal position	
46.	FX398	DUP	X	p22.31	0.36	rs2732615	8,218,944	rs2018496	8,576,027	CNP
47.	FX403	DUP	22	q11.23	0.28	rs3912046	23,967,181	rs4822638	24,249,632	CNP
48.	FX444	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
49.	FX444	DUP	22	q11.23	0.28	rs3912046	23,967,181	rs4822638	24,249,632	CNP
50.	FX444	DEL	X	q26.3	2.3	rs2071718	137,664,347	rs10521805	139,947,192	Verify by gQPCR
51.	FX444	DEL	X	q27.2	2	rs7877128	140,649,041	rs5908815	142,651,304	Verify by gQPCR
52.	FX444	DEL	X	q27.3	1.5	rs2815669	143,463,144	rs5965692	144,985,690	Verify by gQPCR
53.	FX497	DUP	14	q11.2	0.16	rs4412905	19,336,854	rs10141075	19,489,991	CNP
54.	FX497	DEL	15	q11.2	0.98	rs4931862	18,427,103	rs1346662	19,407,629	CNP
55.	FX497	DUP	15	q26.3	0.15	rs7182422	98,752,615	rs8024825	98,905,734	CNP
56.	FX497	DEL	20	p11.21	0.03	rs3746726	22,936,426	rs2424503	22,969,778	CNP
57.	FX509	DEL	11	q11	0.14	rs11823507	55,217,364	rs7951100	55,364,837	CNP
58.	FX509	DUP	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
59.	FX509	DEL	14	q31.3	0.07	rs1088638	86,533,539	rs2301487	86,603,364	No genes in region
60.	FX509	DUP	16	p11.21	0.45	rs1088638	31,010	rs2301487	487,298	CNP
61.	FX509	DEL	20	p11.21	0.03	rs3746726	22,936,426	rs2424503	22,969,778	CNP
62.	FX512	DUP	17	p13.1	0.5	rs1614133	31,429,427	rs2277662	31,923,810	CNP
63.	FX512	DUP	19	p12	0.17	rs8100262	20,613,342	rs587212	20,789,636	Not CNP, XLMR family
64.	FX512	DUP	X	p22.11	1.71	rs17332633	23,174,529	rs5944738	24,937,819	Verify by gQPCR
65.	FX512	DUP	X	p11.3	1.47	rs5952991	43,582,574	45054364	45,054,364	Verify by gQPCR
66.	FX516	DEL	2	q37.3	0.08	rs12623542	242,636,531	rs6759916	242,717,659	CNP
67.	FX516	DEL	15	q11.2	0.27	rs1346662	19,821,421	rs4300610	20,089,383	CNP
68.	FX517	DEL	8	p23.2	0.08	rs1217655	4,293,738	rs7004990	4,376,431	CNP
69.	FX517	DEL	X	q25	0.14	rs488876	129,100,777	rs5932743	129,238,278	Verify by gQPCR

	PATIENT	DUP/ DEL	chromosome	BAND	SIZE (Mb)	START		END		ACTION
						SNP id	Chromosomal position	SNP Id	Chromosomal position	
70.	FX520	DEL	5	q25	0.01	rs1428419	123,563,404	rs922555	123,573,747	No genes in region
71.	FX520	DUP	11	p15.1	0.03	rs1384649	18,896,898	rs7104587	18,930,870	CNP
72.	FX520	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
73.	FX523	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
74.	XMR8	DUP	11	q21	0.04	rs12271216	96,191,471	rs17129651	96,195,684	No genes in region
75.	XMR8	DUP	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
76.	XMR8	DUP	X	q25	0.95	rs12396914	122,318,959	rs5958482	123,271,569	Verify by gQPCR
77.	XMR12	DEL	1	q21.1	1.09	rs7544630	144,607,366	rs0	145,700,996	CNP
78.	XMR12	DEL	4	p16.3	0.18	rs10937909	4,102,877	rs17629551	4,283,407	Verify by gQPCR
79.	XMR12	DUP	7	q11.23	0.58	rs1019096	75,788,276	rs17722395	76,367,046	CNP
80.	XMR12	DEL	9	q21.1	0.02	rs4745105	71,476,355	rs2131039	71,491,754	No genes in region
81.	XMR12	DEL	20	p11.21	0.03	rs3746726	22,936,426	rs2424503	22,969,778	CNP
82.	XMR13	DEL	9	q24.1	0.02	rs10815203	5,374,069	rs1535454	5,400,723	CNP
83.	XMR13	DUP	14	q11.2	0.39	rs2075497	21,634,087	rs17118670	22,029,666	CNP
84.	XMR13	DUP	17	q12	0.05	rs1614133	31,429,427	rs2277662	31,923,810	CNP
85.	XMR14	DEL	3	q25.32	0.25	rs9840667	159,349,247	rs9876322	159,601,940	CNP
86.	XMR14	DEL	8	q24.23	0.17	rs894089	137,757,412	rs7842167	137,931,617	CNP
87.	XMR14	DUP	11	p15.1	0.03	rs1384649	18,896,898	rs7104587	18,930,870	CNP
88.	XMR14	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP
89.	XMR14	DUP	17	p13.1	0.5	rs1614133	31,429,427	rs2277662	31,923,810	CNP
90.	XMR14	DEL	20	p11.21	0.03	rs3746726	22,936,426	rs2424503	22,969,778	CNP
91.	XMR14	DEL	X	q25	0.14	rs488876	129,100,777	rs5932743	129,238,278	Verify by gQPCR
92.	ANG13	DEL	14	q11.2	0.15	rs1782195	19,336,854	rs8022497	19,489,991	CNP

APPENDIX3A: PEDIGREES OF FAMILIES FOR X-CHROMOSOME LINKAGE ANALYSIS

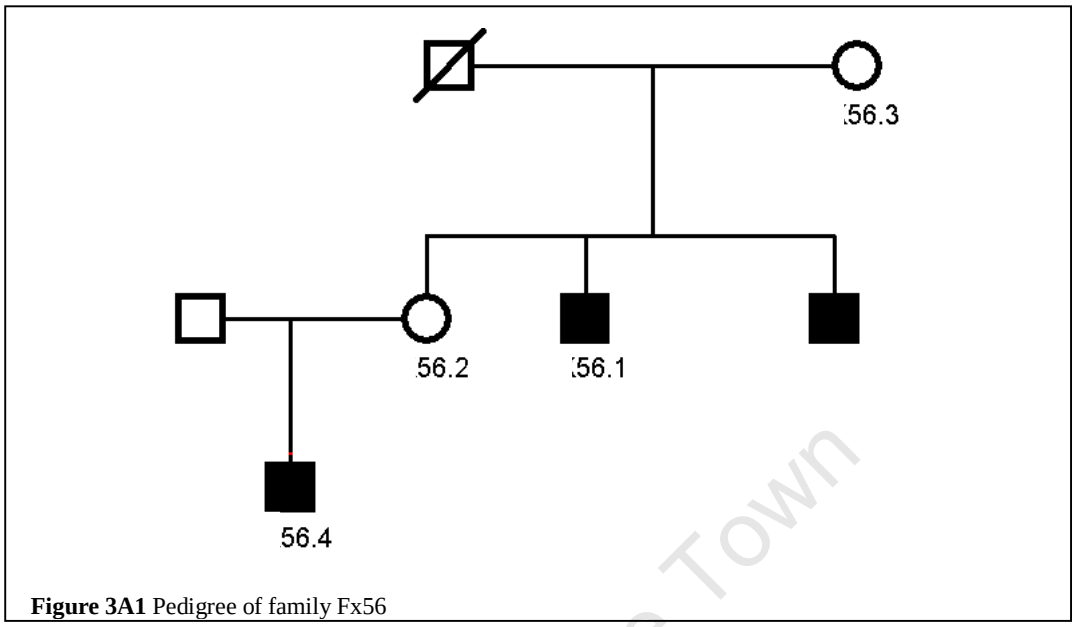


Figure 3A1 Pedigree of family Fx56

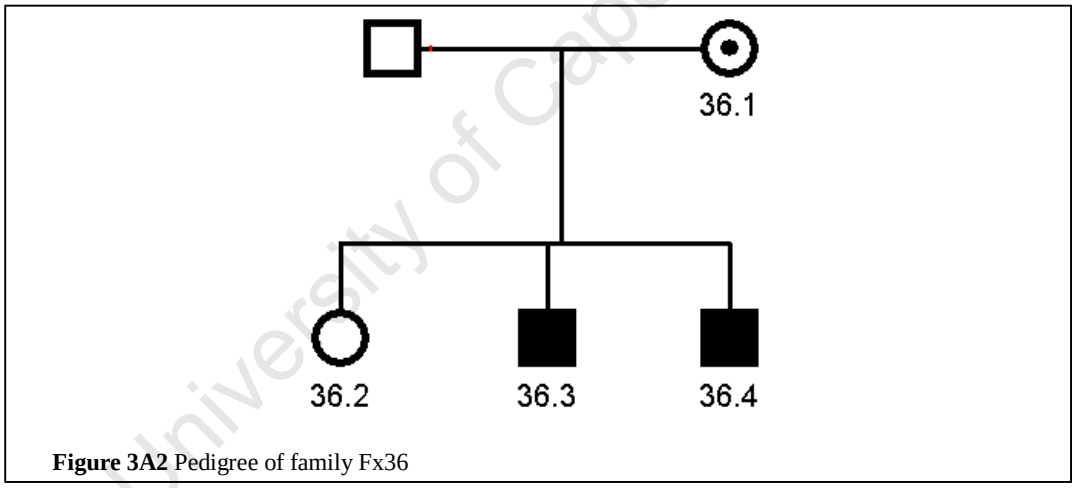


Figure 3A2 Pedigree of family Fx36

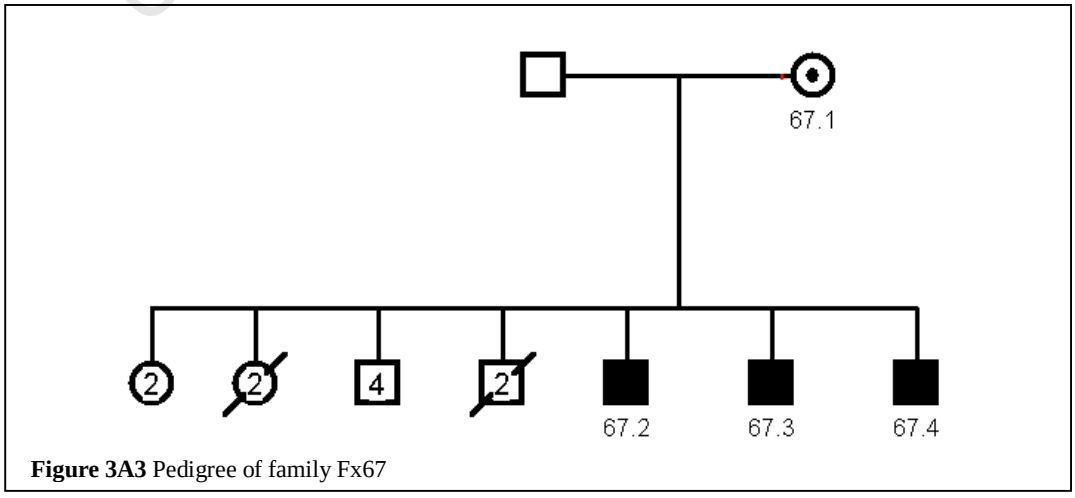


Figure 3A3 Pedigree of family Fx67

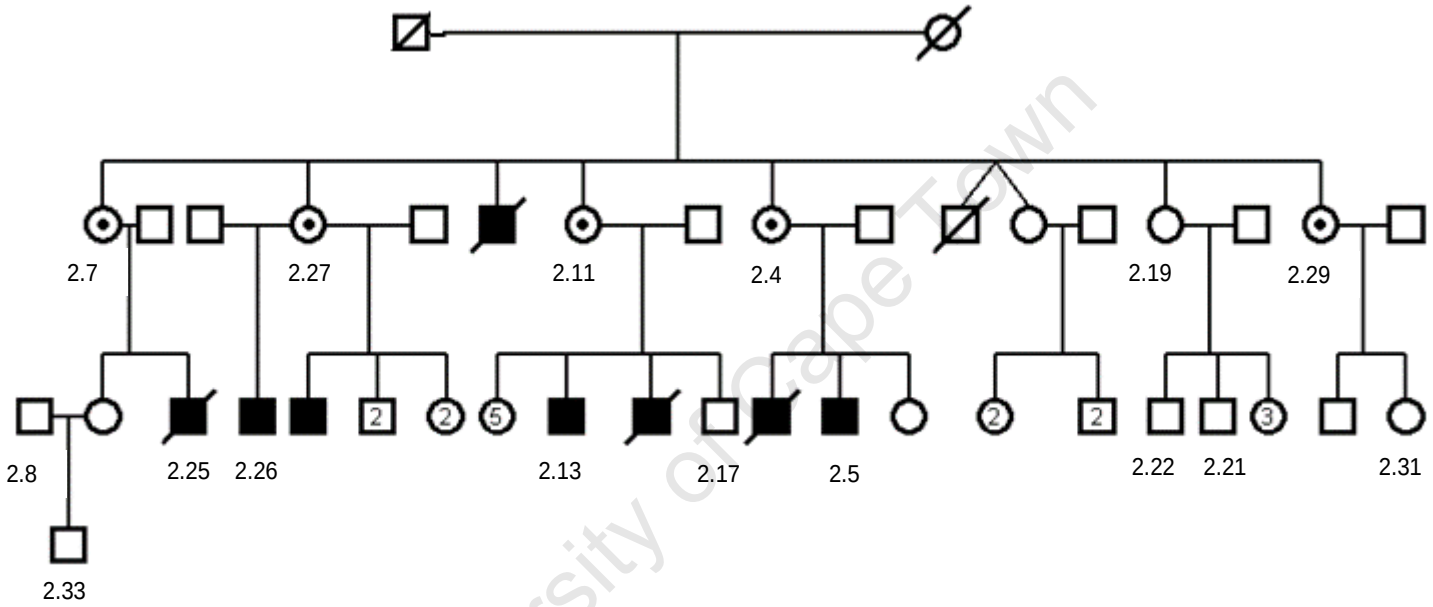


Figure 3A4 Pedigree of family XMR2

APPENDIX3B: MICROSATELLITE MARKERS USED FOR XLMR LINKAGE ANALYSIS.

Marker	Position on X (Mb)	Families genotyped				Hetero-zygosity ratio	Allele size range (bp)	Fluorescent tag
DXS1060	**4.8**			**FX67**		**0.8**	**134-150**	**Ned**
DXS987	**15.7**			**FX67**		**0.83**	**256**	**Fam**
DXS1226	22.8			FX67		0.89	199-223	Ned
DXS1048	26.8			FX67		0.74	109	Unlabelled
DXS1218	29.2			FX67		0.67	255-275	Unlabelled
DXS9896	29.5			FX67		0.82	192-232	Fam
DXS1214	**31**		**FX56**	**FX67**	**XMR2**	**0.79**	**285-298**	**Hex**
DXS1235	31.7			FX67			233	Fam
DXS1237	31.8			FX67			156	Hex
DXS1068	**38**		**FX56**	**FX67**	**XMR2**	**0.79**	**244-264**	**Hex**
DXS8042	39.7			FX67		0.72	145-173	Unlabelled
DXS993	**40**	**FX36**	**FX56**	**FX67**	**XMR2**	**0.79**	**267-293**	**Fam**
DXS6810	42	FX36	FX56	FX67	XMR2	0.69	208-223	Fam
G10578	44.7	FX36	FX56	FX67	XMR2	0.79	218-261	Fam
DXS1199	53.7	FX36	FX56	FX67	XMR2	0.76	277-295	Fam
DXS991	**55**	**FX36**	**FX56**	**FX67**	**XMR2**	**0.8**	**313-341**	**Ned**
DXS1190	55.5	FX36	FX56	FX67	XMR2	0.68	260	Hex
DXS6785	64.4	FX36	FX56	FX67	XMR2	0.77	143-163	Fam
DXS8111	68.3				XMR2	0.72	161-171	Hex
DXS8052	69.6				XMR2	0.78	90-110	Fam
DXS559	70.8	FX36	FX56	FX67	XMR2	0.74	230-242	Unlabelled
DXS441	75.2				XMR2	0.75	173-183	Fam
DXS56	76.8				XMR2	0.75	260-290	Hex
DXS986	**79**	**FX36**	**FX56**	**FX67**	**XMR2**	**0.77**	**151-181**	**Fam**
DXS990	**93**		**FX56**	**FX67**	**XMR2**	**0.74**	**122-132**	**Fam**
DXS6789	95.3			FX67		0.78	118-150	
DXS1106	**102**		**FX56**	**FX67**	**XMR2**	**0.67**	**126-140**	**Hex**
DXS8055	**114**		**FX56**	**FX67**	**XMR2**	**0.65**	**312-324**	**Hex**
DXS1001	**119**		**FX56**	**FX67**	**XMR2**	**0.82**	**191-211**	**Hex**
DXS1047	**128**		**FX56**	**FX67**	**XMR2**	**0.81**	**156-172**	**Hex**
DXS1227	**140**		**FX56**	**FX67**	**XMR2**	**0.73**	**79-99**	**Fam**
DXS8043	**143**		**FX56**		**XMR2**	**0.8**	**146-180**	**Ned**
DXS8091	**147**		**FX56**		**XMR2**	**0.78**	**80-102**	**Hex**

Microsatellite markers of the ABI® Prism Linkage Mapping Set V2 Panel 28 [Applied Biosystems] appear in bold

Fine mapping markers appear in normal (non-bold) font.

APPENDIX3C: GENERULER™ 100BP PLUS DNA LADDER [FERMENTAS]

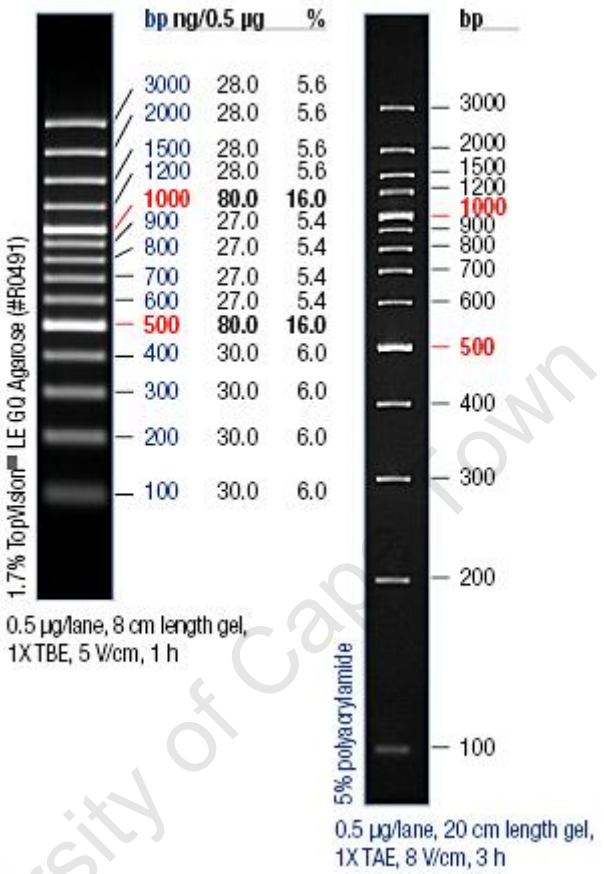


Figure 3C1 Sizes of the bands of the Generuler™ 100bp plus DNA ladder [Fermentas] (as provided by manufacturer at http://www.fermentas.com/en/products/all/dna-electrophoresis/generuler-dna-ladders/sm032

APPENDIX3D: DETAILS OF PCR AMPLIFICATION OF ALL FINE-MAPPING MICROSATELLITE MARKERS USED IN GENOTYPING ANALYSIS FOR XLMR FAMILIES

Table 3D1 PCR reagents used for amplification of the fine mapping microsatellite markers used in the genotyping analysis of XLMR families.

REAGENT (stock concentration)	**MICROSATELLITE MARKER**			
	DXS6810 DXS559	**G10578**	**DXS1199**	**DXS1190/ DXS6785/ DXS411/ DXS8111/ DXS8052/ DXS56/ DXS6789/DXS1235/ DXS1237/DXS1048/ DXS1218/ DXS8042/ DXS9896**
Forward Primer (20µM)	8µM	8µM	10µM	10µM
Reverse Primer (20µM)	8µM	8µM	10µM	10µM
dNTP's (5µM) (Bioline)	5µM	-	-	5µM
Buffer (5X) (Promega)	1×	-	-	1×
Failsafe buffer J (2X) (Epicentre Technologies)	-	1×	1×	-
Go Taq Polymerase (5U/µl) (Promega)	0.5U	0.5U	0.5U	0.5U
DNA	50ng	50ng	50ng	50ng
FINAL REACTION VOLUME	10µl	10µl	10µl	10µl

Table 3D2 PCR conditions for amplification of the fine mapping microsatellite markers used in the genotyping analysis of XLMR families.

Condition	**MICROSATELLITE MARKERS**		
	DXS6810	**G10578/ DXS1199/ DXS1190/ DXS6785**	**DXS559/ DXS411/ DXS8111/ DXS8052/ DXS56/ DXS1235/ DXS1237/ DXS1048/ DXS1218/ DXS8042/ DXS9896**
Initial denaturation	94ºC (3 min)	94ºC (3 min)	95ºC (5 min)
Cycling conditions (30 cycles) Denaturation Primer Annealing Template Elongation	 94ºC (30sec) 55 ºC (75 sec) 72 ºC (15 sec)	 94 ºC (1min) 55 ºC (1min) 72 ºC (1min)	 95 ºC (30 sec) 56 ºC (30 sec) 68 ºC (30 sec)
Final template elongation	72 ºC (5 min)	72 ºC (7 min)	72 ºC (5 min)

APPENDIX4A: PCR PRIMERS USED IN THE AMPLIFICATION OF ALL POSITIONAL CANDIDATE GENES

Table 4A1 Primer pairs used to amplify the *ATRX* mutation hotspots

Primer set	Primer	Tm ('c)	Gc (%)	Mer (bp)	Product(bp)	Sequence (5'-3')
PHD domain	Exon 5 Fwd	50	45	20		attactatgcagagcttgcc
	Exon 9 Rev	53	45	22	691, 581, 467 *	gtgcggaataagagtaggttac
Helicase domain hotspot 1	Exon 17 Fwd	54	55	20	626	tgtgggattgctgctgtgag
	Exon 21 Rev	52	39	23		tctattcctgaactccttaatgg
Helicase domain hotpot 2	Exon 25 Fwd	50	43	21	687	gttgaagtgattaaggtctgg
	Exon 30 Rev	50	45	20		tgcttagttacttgccgatc

* There are three isoforms of the *ATRX* mRNA transcript, the three product sizes (691bp, 581bp, 467bp) correlate with the un-spliced, spliced (exon 7 removed) and a second spliced (exons 6 and 7 removed) isoform respectively.

Table 4A2 PCR primer details for amplification of the nine *ATP6AP2* exons

Primer	Tm ('C)	Gc (%)	Mer (bp)	Product Size (bp)	Sequence (5'-3')
Exon 1 Fwd	53	58	19	360	ctgtcgcttcccaggttac
Exon 1 Rev	52	50	20		atccagagaagatgctgacg
Exon 2 Fwd	54	52	21	333	cgctaagtcagtggtgaatgg
Exon 2 Rev	52	48	21		gtacctacgttcgtttgatgc
Exon 3 Fwd	52	50	20	433	gtcagtgtcatctcagaacc
Exon 3 Rev	53	58	19		gagaacagccaagcctcag
Exon 4 Fwd	50	43	21	402	tgtttaatcttctgtgccctc
Exon 4 Rev	52	50	20		ccaatacccagaggtatgac
Exon 5 Fwd	52	50	20	402	gtcatacctctgggtattgg
Exon 5 Rev	53	43	23		caagtagtataaggaatgggagg
Exon 6 Fwd	54	52	21	248	gtgccttgcttggtaaatggc
Exon 6 Rev	51	41	22		ggatcataaatctagggcatac
Exon 7 Fwd	52	50	20	372	ggctgtgtttgacgtctttc
Exon 7 Rev	52	50	20		ttgaccctactccatccttg
Exon 8 Fwd	52	50	21	364	gtggacttaagccattccac
Exon 8 Rev	51	53	19		ctgctcattcatgtcaccc
Exon 9 Fwd	52	48	21	486	gaatcacagtcactagactgg
Exon 9 Rev	53	58	19		ccacagtgggattcacgtc

Table 4A3 PCR primer details for amplification of the seven *TM4SF2* exons

Primer	Tm ('C)	GC (%)	Mer (bp)	ProductSize (bp)	Sequence (5'-3')
Exon 1 Fwd	59	78	18	428	cccgtgtcgcctcgcctg
Exon 1 Rev	58	65	20		atggacagcgactcggccac
Exon 2 Fwd	55	50	22	350	ccacattgctctgtgctgattc
Exon 2 Rev	54	62	21		tgtgacctcaagacgaaccac
Exon 3 Fwd	55	50	22	199	ccgctgattgaaagtaacctac
Exon 3 Rev	54	52	21		cacccctaactgagttctgc
Exon 4 Fwd	54	55	20	227	tggagtcagtttgggctgtg
Exon 4 Rev	55	50	22		caacctacaggcagtccataag
Exon 5 Fwd	52	50	20	273	acaaggcagtgattgaggag
Exon 5 Rev	52	50	20		ccaaacattcaccactggtc
Exon 6 Fwd	54	55	20	197	tcgagtacacatagcccagc
Exon 6 Rev	54	52	21		cagaggaaaggttagaccagg
Exon 7 Fwd	54	55	20	234	gtgtggttcggtgacttgag
Exon 7 Rev	55	50	22		caagatcaggtagctcaatggc

APPENDIX 5A: ALL *ARX* MUTATIONS REPORTED TO DATE

Table 5A1: ARX mutations associated with non-malformation clinical presentation

Mutation	Protein change	Disorder	# cases		Reference
			Patients	Families	
Expansions of the first Polyalanine tract					
c.298_330dupGCGGCA$(GCG)_9$	A110_A111insAAAAAAAAAA (16 → 25 ALA)	Ohtahara syndrome	2	1	(Kato et al. 2007)
c.304ins$(GCG)_7$	A102_A111insAAAAAAAA (16 → 23 ALA)	ISSX/West syndrome	10	2	(Stromme et al. 2002b)
		West syndrome and dystonia	3	1	(Wohlrab et al. 2005)
		ISSX with severe dyskinetic quadriparesis	6	4	(Guerrini et al. 2007)
		MR and Lennox-Gastaut epilepsy syndrome	1	1	(Wallerstein et al. 2008)
		MRX (including tonic seizures and dystonia but not infantile spasms)	1	1	(Shinozaki et al. 2009)
		MRX (non-epileptic seizures, movement disorder)	1	1	(Poirier et al. 2008)
		Ohtahara syndrome with complex movement disorder	1	1	(Absoud et al. 2009)
c.304ins$(GCG)_2$	A102_A111insAA (16 → 18 ALA)	MRX (moderate MR, no seizures)	2	1	(Bienvenu et al. 2002)
c.333ins$(GCG)_3$	A102_A111insAAA (16 → 19 ALA)	MRX (hypotonia, autism spectrum disorder)	1	1	(Gronskov et al. 2004)
Total expansions of first polyalanine tract			**28**	**14**	

Mutation	Protein change	Disorder	# cases		Reference
			Patients	Families	
Expansions of second polyalanine tract					
c.423_455dup	p.A155_W156 insAAAAAAAAAA (12→22 alanines)	MRX (epilepsy, dystonia, intermittent hyperventilation)	2	1	(Demos et al. 2009)
c.428_451dup	p.A155_W156 insAAAAAAAA (12→20 alanines)	MRX(moderate, no seizures, speech deficit, sporadic-seizures)	14	5families 1 sporadic	(Bienvenu et al. 2002)
		MRX	11	2	(Stromme et al. 2002b)
		Partingtons syndrome	13	2	(Stromme et al. 2002b)
		ISSX/ West syndrome	7	1	(Stromme et al. 2002b)
		MRX	30	4	(Stepp et al. 2005)
		West syndrome	Sporadic		(Kato et al. 2003)
		MRX (including seizures and dystonia)	8	3	(Partington et al. 2004)
		MRX (including Partington sign and hypotonia)	1	1	(Gronskov et al. 2004)
		MRX (including speech difficulties, macroorchidism, lower limb spasticity or foot dystonia and facial telangiectasia)	18	5	(Szczaluba et al. 2006)
		MRX (mild-severe MR, focal hand dystonia, spasticity, seizures, autism, brain cysts, macroorchidism, facial telangiectasia)	21	5	(Nawara et al. 2006)
		Partingtons syndrome, MRX (mild-severe,speech delay, hypotonia)	9	4	(Poirier et al. 2006)
		MRX (moderate – profound and minor congenital anomalies)	5	1	(Laperuta et al. 2007)
		MRX	3	1	(Rujirabanjerd et al. 2007)
		Partingtons syndrome	1	1	(Rujirabanjerd et al. 2007)
		MRX	3	1	(Gestinari-Duarte Rde et al. 2006)
c.430_456dup	p.A155_W156 insAAAAAAAAA (12→21 alanines)	MRX (profound MR, including infantile-onset seizures)	3	1	(Reish et al. 2009)
Total expansions of second polyalanine tract			**150**	**40**	

Mutation	Protein change	Disorder	# cases		Reference
			Patients	Families	
c.81C>G	p.Y27X alternate AUG usage p.M41_C562	Ohtahara syndrome/ West Syndrome	2	1	(Fullston et al. 2010)
c.98T>C	p.L33P	MRX (moderate to profound, sei(Stromme et al. 2002a)(Stromme et al. 2002a)zures)	7	1	(Bienvenu et al. 2002)
c.112C>T	p.P38S	MRX (including speech delay, truncular obesity, autism)	2	1	(Poirier et al. 2006)
c.490A>G	p.Q163R	MRX (moderate to profound MR, includes cerebellar symptoms, seizures)	6	1	(Bienvenu et al. 2002)
c.856G>A	p.G286S	MRX (moderate MR, includes macrocephaly, scoliosis, hypertelorism, seizures)	4	1	(Bienvenu et al. 2002)
c.1058C>T	p.P353L	XMESID	6	1	(Scheffer et al. 2002, Stromme et al. 2002b)
c.1465delG	Fs491X	MRX (affected female, seizures)	1	1	(Wallerstein et al. 2008)
IVS4-816_EX5701 Del (1517bp)	p.R484fs	ISSX/ West syndrome	2	1	(Scheffer et al. 2002)
Total of all on-malformation *ARX* mutation positives			**208**	**62**	

Table 5A2: ARX mutations associated with a malformation clinical presentation

Mutation	Protein change	Disorder	# cases		Reference
			Patients	Families	
c.196+2T>C	Exon1 skipping	XLAG	sporadic		(Kato et al. 2004)
c.232G>T	p.E78X	XLAG (ACC in females)	5 (3 females)	2	(Kato et al. 2004)
c.335_368del	p.A112_G123del&fs	XLAG	Sporadic		(Kato et al. 2004)
c.392_452del	p.P131_A151del&fs	XLAG	Sporadic		(Kato et al. 2004)

c.420_451del	Fs140X	XLAG	1 confirmed (familial)		(Kitamura et al. 2002)
Mutation	**Protein change**	**Disorder**	**# cases**		**Reference**
			Patients	**Families**	
c.428-451dup	p.A155_W156insAAAAAAAA (12→20Ala)	Transsphenoidal encephalocele and hypopituitarism	2	1	(Van Esch et al. 2004)
c.617delG	p.G206fs	XLAG	1 confirmed (familial)		(Kato et al. 2004)
c.619_647del	p.V207_A216del&fs	XLAG	1 confirmed (familial)		(Kato et al. 2004)
c.790delC	p.R264fsX324	XLAG with ACC (Proud syndrome)	sporadic		(Uyanik et al. 2003)
		XLAG	1 confirmed (familial)		(Kitamura et al. 2002)
c.994C>T	p.R332H	XLAG with ACC (Proud syndrome)	sporadic		(Uyanik et al. 2003)
c.995 G>A	p.R332H	XLAG	sporadic		(Kitamura et al. 2002)
c.995G>C	p.R332P	XLAG	sporadic		(Kato et al. 2004)
c.998C>A	p.T333N	Proud/ACC	3	1	(Kato et al. 2004)
c.1028T>A	p.L343Q	XLAG	2	1	(Kitamura et al. 2002)
c.1058C>G	p.P353R	XLAG	1 confirmed (familial)		(Kato et al. 2004)
c.1105G>T	p.E369X	hydraencephaly, abnomal genetalia	sporadic		(Kato et al. 2004)
c.1117C>T	p.Q373X	XLAG	sporadic		(Kitamura et al. 2002)
c.1119+1G>C	Skipping of exon 3	XLAG	1 confirmed (familial)		(Kato et al. 2004)
c.1188insC	Fs396X	XLAG	sporadic		(Kitamura et al. 2002)
c.1372delG	Fs457X	XLAG	sporadic		(Kitamura et al. 2002)
		XLAG	sporadic		(Kato et al. 2004)
c.1561G>A	p.A521T	XLAG/LCH	sporadic		(Kato et al. 2004)
EX1_2Del	No protein	XLAG	1 confirmed (familial)		(Kitamura et al. 2002)
EX2_5Del	G66_C562Del	XLAG (ACC in females)	2	1	(Kato et al. 2004)
Total of malformation *ARX* mutation positives			**34**	**26**	

ACC: Ageneis of Corpus Callosum

Total number of ARX patients: 242 patients (approximately) from 88families.

APPENDIX5B: ALL PUBLISHED MUTATIONS IN *KDM5C*

Table 5B: Mutations in *KDM5C* associated with XLMR

EXON	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	PHENOTYPE	REFERENCE
2	c. 202_203insC	Arg68fsX7	Severe MR No speech Strabismus Abnormal EEG Gallstones Aggressive behaviour	(Jensen et al. 2005)
3	c.229G>A	Ala77Thr	Severe MR Broad hands with brachydactyly Delayed speech Prominent nasal bridge and ears Club foot Aggressive behaviour Affected females (mild MR)	(Abidi et al. 2008)
3	c.260A>G	Asp87Gly	Mild MR No speech	(Tzschach et al. 2006)
8	c.994C>T	Arg332X	Severe MR No speech Mother- mild MR	(Tzschach et al. 2006)
9	c.1162G>C	Ala388Pro	Mild MR Microcephaly Mild dysmorphic facial features	(Jensen et al. 2005)
9	c.1204G>T	Asp402Tyr	Severe MR Short stature No speech Aggressive behaviour	(Jensen et al. 2005)
10	c. 1353C>G	Ser451Arg	Severe MR Mild dysmorphism – large ears Overfriendly, anxious.	(Santos et al. 2006)
11	c.1510G>A	Val504Met	Mild-moderate MR Speech delay Short stature Camptodactyly, clinodactoyly Behavioural problems	(Abidi et al. 2008)
Intron 11	c.1583+5>A	Glu468fsX2	Severe MR Short stature Microcephaly Mild brachydactyly Mild asymmetry of legs and arms Hypertonia brisk plantar reflexes	(Abidi et al. 2008)
12	c.1160C>A	Pro544Thr	Moderate MR Short stature Microcephaly Prominent ears High/prominent nasal bridge Females- learning disabilities/ mild MR	(Rujirabanjerd et al. 2009)
14	c.1924T>C	Phe642Leu	Severe MR No speech Seizures Aggressive behaviour	(Tzschach et al. 2006)

EXON	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	PHENOTYPE	REFERENCE
15	c.2080C>T	Arg694X	MR short stature Strabismus Myopia	(Jensen et al. 2005)
15	c.2092G>A	Glu698Lys	Severe MR	(Jensen et al. 2005)
15	c.2191C>T	Leu731Phe	Severe MR Slow progressive spastic paraplegia Facial hypotonia Maxillary hypoplasia Strabismus, Aggressive behaviour	(Jensen et al. 2005)
15	c.2248C>T	Arg750Trp	Severe MR No speech	(Tzschach et al. 2006)
16	c.2252A>G	Tyr751Cys	Moderate MR	(Tzschach et al. 2006)
16	c.2296C>T	Arg766Trp	Autism	(Adegbola et al. 2008)
21	c.3258_3259insC	Lys1087fsX43	Severe MR Speech delay Short stature Epilepsy Prominent ears, nose, eyebrows Abnormal skin pigmentation Hyperreflexia, spasticity Flexion contractor Broad feet, bulbous fingertips Female-learning disability	(Rujirabanjerd et al. 2009)
23	c.3864G>A	Trp1288X	Severe MR Spasiticity Epileptic seizures Short stature Microcephaly Hypermetropia Affected female with mood disorder	(Jensen et al. 2005)
26	c.4441_4442delAG	Arg1481fsX9	Severe MR Speech delay Macrocephaly Short stature Seizures hyperreflexia	(Abidi et al. 2008)

APPENDIX 5C: PATIENT COHORTS FOR MUTATION DETECTION IN FUNCTIONAL CANDIDATE GENES.

Table 5C1 XLMR families and sib-pairs selected for *ARX* mutation screening

#	FAM#	INHERITANCE	PHENOTYPE	#	FAM#	INHERITANCE	PHENOTYPE
1.	FX 2	Sib-pair*	Female affected	26.	FX89	XLMR?	
2.	FX4	Sib-pair*	Female (twins) affected	27.	FX91	Sib-pair	
3.	FX11	XLMR	Female affected	28.	FX95	Sib-pair*	Long face, large ears
4.	FX15	Sib-pair*	Female affected	29.	FX98	XLMR	Large ears
5.	FX22	XLMR		30.	FX100	XLMR	Large ears
6.	FX25	Sib-pair		31.	FX102	Sib-pair	Large ears, forehead testicles; tall narrow face, prominent chin
7.	FX36	Sib-pair					
8.	FX38	XLMR?		32.	FX105	XLMR	
9.	FX42	XLMR?		33.	FX116	XLMR	Hyperactive, dysmorphic faces
10.	FX45	Sib-pair		34.	FX124	Sib-pair	Large ears
11.	FX46	Sib-pair	Severe MR	35.	FX126	XLMR	Almond shaped ears, epicanthic folds
12.	FX48	Sib-pair		36.	FX128	XLMR	Behavioural difficulties
13.	FX55	XLMR?		37.	FX130	Sib-pairs	Long face and ears, large testicles
14.	FX58	XLMR?		38.	FX131	XLMR	Severe MR, autistic features, poor speech development
15.	FX62	XLMR					
16.	FX67	Sib-pair		39.	FX135	Sib-pair*	Low set ears, hypoplastic third face
17.	FX71	Sib-pair*		40.	FX137	XLMR	Female affected previously diagnosed with Soto's syndrome
18.	FX75	XLMR?					
19.	FX76	XLMR	Epilepsy	41.	FX141	XLMR	N genetalia, high forehead
20.	FX78	XLMR?		42.	FX143	XLMR	Large ears
21.	FX79	XLMR		43.	FX147	Sib-pair*	Affected female, mild MR
22.	FX80	XLMR		44.	FX151	Sib-pair	
23.	FX84	Sib-pair		45.	FX157	XLMR	Moderate MR, mild Dysmorphism
24.	FX87	Sib-pair		46.	FX160	XLMR	
25.	FX88	Sib-pair		47.	FX163	XLMR	Female with deafness

#	FAM#	INHERITANCE	PHENOTYPE	#	FAM#	INHERITANCE	PHENOTYPE
48.	FX165	XLMR		73.	FX293	Sib-pairs	large ears, macrocephaly, severe language delay
49.	FX166	XLMR					
50.	FX167	XLMR					
51.	FX177	XLMR	Mild delay, epicanthic folds, mid face hypoplasia, large ears	74.	FX295	XLMR	
52.	FX181	XLMR		75.	FX299	XLMR?	Affected female
53.	FX188	XLMR	Long face, large ears, clinodactyly	76.	FX300	XLMR?	Affected female
54.	FX196	XLMR?		77.	FX302	XLMR?	Affected female
55.	FX198	Sib-pairs		78.	FX312	XLMR	
56.	FX210	Sib-pairs		79.	FX313	XLMR?	
57.	FX217	XLMR?		80.	FX315	XLMR	Macrocephaly, large ears
58.	FX218	XLMR?	Seizures, cocne facial features, mildly dysmorphic, strabismus	81.	FX317	XLMR	Autistic features
59.	FX219	XLMR?	Affected females, large ears, Macrocephaly, seizures, soft dysmorphic	82.	FX338	XLMR?	
60.	FX220	XLMR?		83.	FX346	Sib-pairs	Hyperactive
61.	FX225	XLMR?		84.	FX348	XLMR	
62.	FX228	Sib-pairs		85.	FX353	XLMR?	Affected female
63.	FX231	XLMR	Mild MR	86.	FX360	Sib-pairs	
64.	FX233	Sib-pairs	Dysmorphic, dev delay, large ears, microcephaly	87.	FX361	Sib-pairs	Speech delay
65.	FX234	Sib-pairs	ADHD, epilepsy	88.	FX362	Sib-pairs	
66.	FX239	Sib-pairs		89.	FX366	XLMR?	Epilepsy
67.	FX243	XLMR		90.	FX376	XLMR	
68.	FX244	XLMR?	Dysmorphic ears	91	FX378	XLMR?	
69.	FX256	XLMR	Dysmorphic	92.	FX382	XLMR?	
70.	FX265	XLMR?	Mother mild mental handicap	93.	FX384	XLMR?	
71.	FX277	XLMR	Dev delay, long face, normal ears and testis	94.	FX391	XLMR	
72.	FX288	XLMR		95.	FX398	XLMR	

#	FAM#	INHERITANCE	PHENOTYPE
96.	FX402	Sib-pairs	
97.	FX403	XLMR	Index: mild global delay, FAM: muscle dystrophy, clumsy
98.	FX413	Sib-pairs	
99.	FX415	XLMR?	
100.	FX417	Sib-pairs	
101.	FX420	XLMR?	
102.	FX421	XLMR?	
103.	FX425	XLMR?	Mild Dysmorphism, behavioural deterioration
104.	FX428	XLMR	
105.	FX430	XLMR?	
106.	FX431	XLMR?	
107.	FX434	Sib-pairs	
108.	FX442	XLMR	
109.	FX443	XLMR?	
110.	FX446	XLMR	Global dev delay, hypotonia, mild dysmorphism
111.	FX449	XLMR?	
112.	FX452	XLMR?	
113.	XMR3.1	XLMR	

Sib-pair* = brother-sister pair

Table 5C2 Isolated male MR cases selected for screening of the c. c.428_451dup mutation

#	FAM #	PHENOTYPE	#	FAM #	PHENOTYPE
1.	FX8		24.	FX97	Severe MR
2.	FX12		25.	FX99	anti-mongolian slant, low set ears, preaxial polydactyly – big toes,
3.	FX14				
4.	FX16		26.	FX101	
5.	FX17		27.	FX103	
6.	FX18		28.	FX106	Strabismus
7.	FX19	maternal uncle aggressive	29.	FX107	NF, epilepsy, facial features of FX
8.	FX23		30.	FX108	
9.	FX24	Autism	31.	FX109	
10.	FX26		32.	FX110	
11.	FX28		33.	FX112	epilepsy, macro-orchidism at birth, epicanthic folds
12.	FX29		34.	FX113	mild MR, large jaw and head
13.	FX33	Epilepsy, diplegia,birth asphyxia	35.	FX114	large ears and penis
14.	FX52	tall, long narrow facies, crowded teeth, narrow, high arched palate, long neck, chest wide and narrow, mild webbing of fingers, large testes	36.	FX115	
15.	FX57	socialisation and speech delay, long face, long low set ears, large testicles, brisk reflexes, slight antimongolian slant to papebral fissures	37.	FX117	
16.	FX59	course features, long faceand thick supraorbital ridge and synophris, large testicles, forceps delivery	38.	FX118	
17.	FX70		39.	FX119	
18.	FX72		40.	FX121	
19.	FX73		41.	FX123	
20.	FX82		42.	FX129	behaviour disorder, epilepsy, hyperactivity
21.	FX85	large ears, hyperactive, Prominent forehead, big hand and feet	43.	FX132	
22.	FX90	Epilepsy	44.	FX133	large ears,
23.	FX96	large ears and jaw	45.	FX134	Microcephaly

#	FAM #	PHENOTYPE	#	FAM #	PHENOTYPE
46.	FX138		75.	FX211	FX type dysmorphology
47.	FX139	severe MR, autism, dysmorphic, large ears	76.	FX212	dysmorphology: long face, large ears, chin and testis
48.	FX140		77.	FX213	
49.	FX142		78.	FX214	moderate MR, tall, occasional aggression
50.	FX144	epilepsy	79.	FX216	
51.	FX145	mild MR, large ears	80.	FX221	dysmorphic features of FX
52.	FX146	behavioural problems, long face, ears and large testis	81.	FX222	deafness
53.	FX150	large testicles	82.	FX223	large testicles
54.	FX152		83.	FX224	adopted, autistic, severe anxiety
55.	FX159		84.	FX227	moderate to mild MR, unusual facies with hyperteliarism, back slanting ears, high arched palate, slightly small head circum
56.	FX161	v large ears, microcephalic, tall, **deafness**, impaired vision			
57.	FX168		85.	FX229	severe MR, arthrogryposis, greying hair
58.	FX169	aggressive outbursts	86.	FX230	
59.	FX172	behaviour disorder, long face, testis	87.	FX236	progressive neurodevelopment delay from 2 yrs, ataxia, global speech delay, N MRI
60.	FX174	brilliant blue eyes			
61.	FX175	dev and speech delay, prominent lips, maxillars hypoplasia	88.	FX237	
62.	FX176	macroplasia	89.	FX238	
63.	FX178		90.	FX241	
64.	FX180	big ears and genetalia	91.	FX245	
65.	FX182	large ears, short stature, aggressive behaviour	92.	FX247	dev delay, dysmorphic
66.	FX185		93.	FX248	
67.	FX187		94.	FX249	
68.	FX189		95.	FX250	
69.	FX191	mild dysmorphic features, thin face, min communication	96.	FX252	Hyperactive,seizures
70.	FX197	pervasive developmental disorder (**autism** spectrum)	97.	FX254	ADHD
71.	FX200	developmental delay, social problems, corneal opacity	98.	FX255	
72.	FX206		99.	FX263	
73.	FX207	speech delay	100.	FX271	behavioural problems, mild dysmorphology
74.	FX208	mild developmental delay	101.	FX272	

#	FAM #	PHENOTYPE	#	FAM #	PHENOTYPE
102.	FX274		131.	FX343	small penis, behaviour, hyperplegia, high pain tolerance, dev delay
103.	FX275	large ears	132.	FX345	
104.	FX276	learning disability	133.	FX349	dysmorphic dev delay
105.	FX279	dysmorphic	134.	FX356	
106.	FX282		135.	FX357	
107.	FX289	severe MR,	136.	FX358	global dev delay
108.	FX290	global delay, **autism**, profoundly **deaf**, large ears	137.	FX359	
109.	FX291		138.	FX363	severe MR, autistic features
110.	FX292	dysmorphic, **autistic** behaviour	139.	FX367	
111.	FX297		140.	FX369	severe global delay, autistic features
112.	FX298	Soto's syndrome?	141.	FX370	dev delay
113.	FX305		142.	FX371	
114.	FX306		143.	FX372	global dev delay
115.	FX307	macro-orchidism, prominent jaw	144.	FX381	
116.	FX308		145.	FX386	
117.	FX309		146.	FX387	
118.	FX310		147.	FX389	
119.	FX311		148.	FX390	
120.	FX316		149.	FX395	global delay
121.	FX322		150.	FX396	twins
122.	FX323		151.	FX397	
123.	FX324	short attention span, big ears	152.	FX400	global delay
124.	FX327	pervasive dev disorder,mild dysmorphism	153.	FX401	
125.	FX328	Speech problems, enuresis	154.	FX404	macrocrania, bat ears
126.	FX330	dev delay, behaviour problems, hyperphagia	155.	FX405	
127.	FX331	physical disability	156.	FX406	global delay
128.	FX332	large ears	157.	FX407	marphanoid habitus
129.	FX341	large ears	158.	FX410	
130.	FX342		159.	FX412	autistic behavioural difficulties

#	FAM #	PHENOTYPE
160.	FX416	**autistic** features
161.	FX422	Hydrocephalus, epilepsy, large ears, speech
162.	FX423	
163.	FX427	
164.	FX429	global delay
165.	FX432	
166.	FX439	
167.	FX440	
168.	FX441	
169.	FX445	behavioural difficulties
170.	FX451	
171.	FX453	
172.	FX454	
173.	FX455	physical disability
174.	FX458	
175.	FX459	
176.	FX461	
177.	FX462	Global dev delay
178.	FX463	
179.	FX472	
180.	FX477	Prognathism, big ears, macrocephaly
181.	FX478	Schizophrenic
182.	FX482	
183.	FX483	

Table 5C3 XLMR patients selected for *KDM5C* mutation detection

#	FAM #	PHENOTYPE
1.	FX46	severe MR
2.	FX76	epilepsy
3.	FX98	Large ears
4.	FX100	Moderate MR, large ears
5.	FX102	large ears, forehead and testicles, tall narrow face, prominent chin(SIB PAIRS)
6.	FX116	hyperactivity, dysmorphic facies
7.	FX124	large ears (SIB PAIRS)
8.	FX128	behavioural difficulties
9.	FX131	severe MR, autistic features, poor speech development
10.	FX135	low set ears and hypoplastic central third face (SIB PAIRS)
11.	FX143	Large ears
12.	FX148	Moderate MR
13.	FX157	moderate MR, mild dysmorphism
14.	FX177	mild delay, epicanthic folds, mid face hypoplasia, large ears
15.	FX188	long face, large ears, webbing of fingers, clinodactyly
16.	FX233	dysmorphic, dev delay, large ears, microcephaly
17.	FX256	dysmorphism
18.	FX277	dev delay, long face, normal ears and testis, epilepsy
19.	FX293	large ears, macrocephaly, severe language delay (SIB PAIRS)
20.	FX317	autistic features
21.	FX361	speech delay
22.	FX373	global delay, behaviour problems
23.	FX56	Linkage
24.	FX36	Linkage
25.	XMR2	Severe MR spastic paraplegia, linkage

APPENDIX 5D: PCR PRIMERS USED IN THE AMPLIFCATION OF ALL FUNCTIONAL CANDIDATE GENES.

Table 5D.1 Primer sets used to amplify the five *ARX* exons for downstream mutation detection.

Exon	Primer	amplicon size (bp)	Mer(bp)	GC	Tm(°C)	Sequence (5' – 3')
1	Fwd	475	20	45	58	tccgttataacccgctatct
1	Rev	475	19	53	58	acagccctggctagatgtt
2A	Fwd	658	18	61	58	ccaaggcgtcgaagtctg
2A	Rev	658	20	50	60	tcatcttcttcgtcctccag
2B	Fwd	528	18	72	62	accggcaccgaggacgac
2B	Rev	528	20	65	64	gagtccaggagccaagcgtc
3	fwd	262	19	58	60	agtaggcctgccatagagg
3	Rev	262	20	45	58	tgatcctgcttctcttggtt
4	Fwd	777	20	45	58	ttgaagttgcggctcctatt
4	Rev	777	20	50	60	ggttgtcacggttgtcgtta
5	Fwd	573	18	61	58	caggaaagccctctctgc
5	Rev	573	20	55	54	gcatccagactgctgtgaag

Table 5D.2 Primer sets used to amplify the 26 *JARID1C* exons for downstream mutation detection

Exon	Primer	amplicon size (bp)	Mer(bp)	GC	Tm(°C)	Sequence (5' – 3')
1	Fwd	526	54	52	21	gacgctgacaaaccaagatgg
1	Rev	526	54	52	21	cctactgcttcattccgtctc
2	Fwd	175	56	44	25	cactatgctgagataactcaagtcc
2	Rev	175	55	48	23	ctcaggtatacattctcccaacc
3	Fwd	288	52	50	20	ccaaagatagtggtcagtgg
3	Rev	288	52	48	21	ccttccattgcagcctaaaga
4	fwd	294	51	53	19	ctttctacaggcctactcc
4	Rev	294	52	48	21	atctgtgctgaagggtaaagc
5	Fwd	249	56	60	20	ctggagtccatgtcctgacc
5	Rev	249	54	55	20	agccttagccagaaggaagg
6 and 7	Fwd	533	56	60	20	gtctgccccagatagcagtc
6 and 7	Rev	533	56	60	20	cgaactggtccagtgccatg
8	Fwd	349	55	48	23	gacctagcatgactagcctatac
8	Rev	349	57	55	22	gactatggctgagctaagaggc
9 and 10	Fwd	601	54	52	21	cagttccacttgggaggattc
9 and 10	Rev	601	55	50	22	ctcatggctacataagacaggc
11 and 12	Fwd	530	54	52	21	cttagcataaccctcatgccc
11 and 12	Rev	530	54	55	20	aatcactcctgccgcttgtc
13 and 14	fwd	594	52	50	20	gtctgggattctgttgtcag
13 and 14	Rev	594	54	55	20	ccaccagaatagggtgcttg
15 and 16	Fwd	533	55	50	22	gaatctaaagtaggggtcggtg
15 and 16	Rev	533	54	52	21	gggaatagaacttgcctgtgg
17	Fwd	296	54	52	21	ccacaggcaagttctattccc
17	Rev	296	54	52	21	ctggatcctcagcaccttatg
18	Fwd	190	54	55	20	aggcactgagtttggacctg
18	Rev	190	53	58	19	gtcccctgatccctcatc
19	Fwd	544	56	57	21	ggtgggacaaggttccatctg
19	Rev	544	56	57	21	gagcgtgatacctaaggccac
20	Fwd	277	56	57	21	gcagaccacatcagactgagc
20	Rev	277	55	63	19	ccaacccatcccagcaacc
21 and 22	fwd	565	52	50	20	tggcaagttgaactgagctg
21 and 22	Rev	565	52	48	21	tatcatcaccaagcccttctc
23	Fwd	716	56	57	21	gtaggctgctgacccactttg
23	Rev	716	55	63	19	cagggccgagcctaaactg
24 and 25	Fwd	503	54	52	21	cttcacttcagttgcccctac
24 and 25	Rev	503	55	63	19	tactcggcctgacctcctg
26	Fwd	527	55	63	19	aggaggtcaggccgagtag
26	Rev	527	54	52	21	caagaagcaggcttgatggtc

APPENDIX 5E: PCR CONDITIONS FOR ARX AMPLIFCATION.

Table 5E1 Reagents utilised in the PCR amplification of 6 amplicons for *ARX*

REAGENT (STOCK CONCENTRATION)	EXON					
	1	**2A**	**2B**	**3**	**4**	**5**
Forward Primer (20µM)	15µM	10µM	40µM	15µM	15µM	10µM
Reverse Primer (15µM)	15µM	10µM	40µM	15µM	15µM	10µM
dNTP's (5µM)[Bioline]	5µM	-	20µM	5µM	5µM	-
Buffer (10X) [Invitrogen]	1×	-	1×	1×	1×	-
Failsafe buffer J (2X) [Epicentre Technologies]	1×	1×	1×	1×	1×	1×
Taq Polymerase (5U/µl) [Invitrogen]	1U	-	2U	1U	1U	-
ELT polymerase (5U/µl) [Roche]	-	2,5U	-	-	-	1,5U
$MgCl_2$ [Invitrogen]	75µM	-	150µM	75µM	75µM	-
DNA	100ng	100ng	400ng	100ng	100ng	100ng
DMSO [Merck]	-	-	8%	-	2,5%	-
FINAL REACTION VOLUME	50µl	25µl	100µl	50µl	50µl	25µl

Table 5E2 Optimised cycling conditions for PCR amplification of 6 *ARX* amplicons

Condition	**Exon**					
	1	2a	2b	3	4	5
Thermal cycler *	A.B	A.B	PX2	A.B	A.B	PX2
Initial denaturation	95ºC (3min)	95ºC(5min)	95ºC(5min)	95ºC(5min)	95ºC(5min)	95ºC(5min)
Cycling conditions (10 cycles) Denaturation Primer Annealing Elongation	95 ºC(30s) 59–54 ºC (30s) 72 ºC(30s)			95 ºC(30s) 59–55 ºC(30s) 68 ºC(30s)		95 ºC(30s) 59–55 ºC (30s) 68 ºC (30s)
Cycling conditions (30 cycles) Denaturation Primer Annealing Elongation	95ºC (30s) 54 ºC (30s) 72 ºC(30s)	95 ºC(1min) 57 ºC(1min) 72 ºC (1min)	95 ºC (1min) 57 ºC (1min) 72 ºC(1min)	95 ºC(1min) 55 ºC (1min) 68 ºC (1min)	95 ºC (1min) 57 ºC (1min) 68 ºC (1min)	95 ºC (1min) 55 ºC (1min) 68 ºC (1min)
Final template elongation	72 ºC (5min)	72 ºC (7min)	72 ºC (7min)	68 ºC (7min)	68 ºC (7min)	68 ºC (7min)

*Thermal cyclers: A.B: GeneAmp® PCR system 9700 [Applied Biosystems]
PX2: Px2 thermal cycler [Thermo Electron Corporation]

APPENDIX 6A: PCR PRIMERS USED TO VALIDATE POTENTIAL DMRS

	Probe ID	Primer	Forward (5’ – 3’)	Reverse (5’ – 3’)
	CHR16FS010387819	Chr16_ATF71P2_Ex1_BSP	ttagttggttgtgtttgattggt	aacttttacttacccctcaaaac
	CHR11FS126375528	CHR11_KIRREL_EX1_BSP	tattttgattgaagaggaagaagta	aacacaacctttttcccccactatc
	CHR17FS034606720	CHR17_CACNB1_EX1_BSP	gaaggtttgttggtatttggta	cccatccttacaaattataattaa
	CHR05FS001853381	CHR5_MRPL36_5’UTR_BSP	agtggataggagaaatgtaaataggtt	cccaaaaatatcccaacaaa
	CHR02FS065138632	CHR2_CEP68_INT1_BSP	tggaaagatggagattagtgatg	aaacccaaaacaaactaaccc
	CHR19FS015436362	CHR19_ZIM_PEG_BSP	gatggtatttaatgggtggg	acaccaatactatccctattaccac
	CHR18FS031177845	CHR18_ZNF24_EX1_BSP	gtgggaaatggagggaagg	aaactctctaacccccctatatcc
	CHR05FS074359693	CHR5_GCNT4_5’UTR_BSP	aagagataagtggagtgttggttat	ccaaaattccatacacaaaaaa

APPENDIX 6B: THE 112 UNIQUE DMRS IDENTIFIED BY MEDIP-CHIP AND STATISITCAL ANALYSES

chromosome	Start position	End position	Chromosome	Start position	End position
X	16713584	16715711	5	131624135	131626245
X	128942083	128944357	5	137700566	137702661
1	25127455	25129575	5	150516390	150518478
1	28112640	28114750	5	174082407	174084517
1	44643249	44645350	5	175156626	175158726
1	54726407	54728517	5	176784401	176787011
1	70592010	70594090	6	31740492	31742702
1	93417417	93419632	6	89846345	89848425
1	109557476	109559581	6	134315526	134318204
1	144183679	144186069	6	139735765	139737840
1	151500233	151502418	7	74825442	74827788
1	152437180	152439156	8	38972194	38974279
1	152442735	152444350	8	97575058	97577188
1	153242508	153244618	8	109163371	109165476
1	153374591	153377005	9	101900242	101902337
1	154964148	154966253	9	129717966	129720151
1	181870746	181872941	10	22649207	22651531
1	204009878	204012050	10	102810609	102813194
2	11804805	11807716	10	119123054	119125254
2	28469900	28472413	11	13941342	13943542
2	44441092	44443200	11	34030116	34032226
2	65137130	65139632	11	57120910	57123428
2	172674115	172676449	11	57883019	57885535
2	175740043	175742263	11	59334117	59336807
2	220083073	220085143	11	62069543	62071728
2	232279156	232281456	11	118393757	118395857
3	49881048	49883447	11	118476701	118478796
3	99797733	99799857	11	126374716	126377220
3	168935011	168937116	12	26169141	26171531
4	30330137	30332543	12	37585584	37587710
4	38481135	38483735	12	52179951	52182039
4	103967233	103969433	12	54995331	54998056
4	115738669	115740854	12	70342897	70345267
5	1852571	1854881	12	78851846	78853956
5	74233023	74234523	12	84197019	84199146
5	74359003	74360988	14	51603918	51606023
5	112100588	112102983	14	63078931	63081128

chromosome	Start position	End position	Chromosome	Start position	End position
14	68328161	68330437	17	35389589	35391689
14	99016701	99018911	17	35508602	35510829
15	41871192	41873507	17	45139582	45141812
15	64780655	64783027	18	31176844	31179350
15	80123746	80125831	18	72336489	72338668
16	10387119	10389529	19	9111291	9113789
16	19802650	19804846	19	11391713	11393834
16	29982732	29984839	19	13086497	13089317
16	30576341	30578526	19	15435262	15437667
16	33869722	33871932	19	38806836	38808336
16	49138052	49140542	19	38859806	38861781
16	56983125	56985295	19	51081155	51083264
16	66677187	66679500	19	53585545	53587610
17	7699038	7701328	19	62041555	62044363
17	27617519	27619810	20	41250092	41252389
17	30593413	30595608	20	43995825	43997940
17	32366016	32368346	22	19113293	19115393
17	33971876	33973972	22	22683333	22684833
17	34605635	34608035	22	22702422	22704546

REFERENCES

Aapola U, Kawasaki K, Scott HS, Ollila J, Vihinen M, Heino M et al. Isolation and initial characterization of a novel zinc finger gene, *DNMT3L*, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. Genomics 2000;65;293-8.

Abidi F, Schwartz CE, Carpenter NJ, Villard L, Fontes M, Curtis M. Carpenter-waziri syndrome results from a mutation in *XNP*. Am J Med Genet 1999;85;249-51.

Abidi FE, Cardoso C, Lossi AM, Lowry RB, Depetris D, Mattei MG et al. Mutation in the 5' alternatively spliced region of the *XNP/ATR-X* gene causes Chudley-Lowry syndrome. Eur J Hum Genet 2005;13;176-83.

Abidi FE, Holinski-Feder E, Rittinger O, Kooy F, Lubs HA, Stevenson RE et al. A novel 2 bp deletion in the *TM4SF2* gene is associated with MRX58. J Med Genet 2002;39;430-3.

Abidi FE, Holloway L, Moore CA, Weaver DD, Simensen RJ, Stevenson RE et al. Mutations in *JARID1C* are associated with X-linked mental retardation, short stature and hyperreflexia. J Med Genet 2008;45;787-93.

Absoud M, Parr JR, Halliday D, Pretorius P, Zaiwalla Z, Jayawant S. A novel ARX phenotype: Rapid neurodegeneration with Ohtahara syndrome and a dyskinetic movement disorder. Dev Med Child Neurol 2009.

Adams JP, Sweatt JD. Molecular psychology: Roles for the ERK MAP kinase cascade in memory. Annu Rev Pharmacol Toxicol 2002;42;135-63.

Adegbola A, Gao H, Sommer S, Browning M. A novel mutation in *JARID1C*/*SMCX* in a patient with autism spectrum disorder (ASD). Am J Med Genet A 2008;146A;505-11.

Agulnik AI, Mitchell MJ, Mattei MG, Borsani G, Avner PA, Lerner JL et al. A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. Hum Mol Genet 1994;3;879-84.

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA et al. Deletion of ultraconserved elements yields viable mice. PLoS Biol 2007;5;e234.

Aicardi J. The etiology of developmental delay. Semin Pediatr Neurol 1998;5;15-20.

Alberts B. Molecular biology of the cell. 5th edition. New York: Garland Science; 2008.

Albrecht A, Mundlos S. The other trinucleotide repeat: Polyalanine expansion disorders. Curr Opin Genet Dev 2005;15;285-93.

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet 2009;41;1061-7.

American Psychiatric Association. Task Force on DSM-IV. Diagnostic and statistical manual of mental disorders : DSM-IV. 4th edition. Washington, DC: American Psychiatric Association; 1994.

Amiel J, Trochet D, Clement-Ziza M, Munnich A, Lyonnet S. Polyalanine expansions in human. Hum Mol Genet 2004;13 Spec No 2;R235-43.

Antar LN, Afroz R, Dictenberg JB, Carroll RC, Bassell GJ. Metabotropic glutamate receptor activation regulates fragile x mental retardation protein and *FMR1* mRNA localization differentially in dendrites and at synapses. J Neurosci 2004;24;2648-55.

Argentaro A, Yang JC, Chapman L, Kowalczyk MS, Gibbons RJ, Higgs DR et al. Structural consequences of disease-causing mutations in the ATRX-DNMT3-DNMT3L (ADD) domain of the chromatin-associated protein *ATRX*. Proc Natl Acad Sci U S A 2007;104;11939-44.

Aruga J, Mikoshiba K. Identification and characterization of *SLITRK*, a novel neuronal transmembrane protein family controlling neurite outgrowth. Mol Cell Neurosci 2003;24;117-29.

Aruga J, Yokota N, Mikoshiba K. Human *SLITRK* family genes: Genomic organization and expression profiling in normal brain and brain tumor tissue. Gene 2003;315;87-94.

Aston E, Whitby H, Maxwell T, Glaus N, Cowley B, Lowry D et al. Comparison of targeted and whole genome analysis of postnatal specimens using a commercially available array based comparative genomic hybridisation (aCGH) microarray platform. J Med Genet 2008;45;268-74.

Aznarez I, Zielenski J, Rommens JM, Blencowe BJ, Tsui LC. Exon skipping through the creation of a putative exonic splicing silencer as a consequence of the cystic fibrosis mutation R553X. J Med Genet 2007;44;341-6.

Badens C, Lacoste C, Philip N, Martini N, Courrier S, Giuliano F et al. Mutations in PHD-like domain of the *ATRX* gene correlate with severe psychomotor impairment and severe urogenital abnormalities in patients with ATRX syndrome. Clin Genet 2006;70;57-62.

Ballas N, Mandel G. The many faces of *REST* oversee epigenetic programming of neuronal genes. Curr Opin Neurobiol 2005;15;500-6.

Ballif BC, Rorem EA, Sundin K, Lincicum M, Gaskin S, Coppinger J et al. Detection of low-level mosaicism by array CGH in routine diagnostic specimens. Am J Med Genet A 2006;140;2757-67.

Barel O, Shalev SA, Ofir R, Cohen A, Zlotogora J, Shorer Z et al. Maternally inherited Birk Barel mental retardation dysmorphism syndrome caused by a mutation in the genomically imprinted potassium channel *KCNK9*. Am J Hum Genet 2008;83;193-9.

Basel-Vanagaite L, Attia R, Yahav M, Ferland RJ, Anteki L, Walsh CA et al. The *CC2D1A*, a member of a new gene family with C2 domains, is involved in autosomal recessive non-syndromic mental retardation. J Med Genet 2006;43;203-10.

Bashiardes S, Kousoulidou L, van Bokhoven H, Ropers HH, Chelly J, Moraine C et al. A new chromosome X exon-specific microarray platform for screening of patients with X-linked disorders. J Mol Diagn 2009;11;562-8.

Baumann C, De La Fuente R. ATRX marks the inactive X chromosome (Xi) in somatic cells and during imprinted X chromosome inactivation in trophoblast stem cells. Chromosoma 2009;118;209-22.

Baumann C, Schmidtmann A, Muegge K, De La Fuente R. Association of *ATRX* with pericentric heterochromatin and the Y chromosome of neonatal mouse spermatogonia. BMC Mol Biol 2008;9;29.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS et al. Ultraconserved elements in the human genome. Science 2004;304;1321-5.

Bejjani BA, Shaffer LG. Application of array-based comparative genomic hybridization to clinical diagnostics. J Mol Diagn 2006;8;528-33.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Statist Soc B 1995;57;289-300.

Bergmann C, Zerres K, Senderek J, Rudnik-Schoneborn S, Eggermann T, Hausler M et al. Oligophrenin 1 (*OPHN1*) gene mutation causes syndromic X-linked mental retardation with epilepsy, rostral ventricular enlargement and cerebellar hypoplasia. Brain 2003;126;1537-44.

Berry-Kravis E, Sumis A, Hervey C, Nelson M, Porges SW, Weng N et al. Open-label treatment trial of lithium to target the underlying defect in fragile X syndrome. J Dev Behav Pediatr 2008;29;293-302.

Berube NG, Healy J, Medina CF, Wu S, Hodgson T, Jagla M et al. Patient mutations alter ATRX targeting to PML nuclear bodies. Eur J Hum Genet 2008;16;192-201.

Berube NG, Jagla M, Smeenk C, De Repentigny Y, Kothary R, Picketts DJ. Neurodevelopmental defects resulting from *ATRX* overexpression in transgenic mice. Hum Mol Genet 2002;11;253-61.

Berube NG, Mangelsdorf M, Jagla M, Vanderluit J, Garrick D, Gibbons RJ et al. The chromatin-remodeling protein ATRX is critical for neuronal survival during corticogenesis. J Clin Invest 2005;115;258-67.

Berube NG, Smeenk CA, Picketts DJ. Cell cycle-dependent phosphorylation of the ATRX protein correlates with changes in nuclear matrix and chromatin association. Hum Mol Genet 2000;9;539-47.

Bhat SS, Rogers RC, Holden KR, Srivastava AK. A novel in-frame deletion in *ARX* is associated with lissencephaly with absent corpus callosum and hypoplastic genitalia. Am J Med Genet A 2005;138;70-2.

Bienvenu T, Chelly J. Molecular genetics of Rett syndrome: When DNA methylation goes unrecognized. Nat Rev Genet 2006;7;415-26.

Bienvenu T, Poirier K, Friocourt G, Bahi N, Beaumont D, Fauchereau F et al. *ARX*, a novel prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation. Hum Mol Genet 2002;11;981-91.

Bienz M. The PHD finger, a nuclear protein-interaction domain. Trends Biochem Sci 2006;31;35-40.

Bijlsma EK, Gijsbers AC, Schuurs-Hoeijmakers JH, van Haeringen A, Fransen van de Putte DE, Anderlid BM et al. Extending the phenotype of recurrent rearrangements of 16p11.2: Deletions in mentally retarded patients without autism and in normal individuals. Eur J Med Genet 2009;52;77-87.

Bird AP. CpG-rich islands and the function of DNA methylation. Nature 1986;321;209-13.

Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T. BiQ analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. Bioinformatics 2005;21;4067-8.

Boichard A, Venet L, Naas T, Boutron A, Chevret L, de Baulny HO et al. Two silent substitutions in the *PDHA1* gene cause exon 5 skipping by disruption of a putative exonic splicing enhancer. Mol Genet Metab 2008;93;323-30.

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19;185-93.

Bonnet C, Leheup B, Beri M, Philippe C, Gregoire MJ, Jonveaux P. Aberrant *GRIA3* transcripts with multi-exon duplications in a family with X-linked mental retardation. Am J Med Genet A 2009;149A;1280-9.

Boudewijns M, van Dongen JJ, Langerak AW. The human androgen receptor X-chromosome inactivation assay for clonality diagnostics of natural killer cell proliferations. J Mol Diagn 2007;9;337-44.

Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. *DNMT3L* and the establishment of maternal genomic imprints. Science 2001;294;2536-9.

Boyes J, Bird A. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 1991;64;1123-34.

Bozon B, Kelly A, Josselyn SA, Silva AJ, Davis S, Laroche S. MAPK, CREB and ZIF268 are all required for the consolidation of recognition memory. Philos Trans R Soc Lond B Biol Sci 2003;358;805-14.

Brown L, Paraso M, Arkell R, Brown S. In vitro analysis of partial loss-of-function *ZIC2* mutations in holoprosencephaly: Alanine tract expansion modulates DNA binding and transactivation. Hum Mol Genet 2005;14;411-20.

Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by taq DNA polymerase: Primer modifications that facilitate genotyping. BioTechniques 1996;20;1004,6, 1008-10.

Cantagrel V, Lossi AM, Boulanger S, Depetris D, Mattei MG, Gecz J et al. Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males. J Med Genet 2004;41;736-42.

Cardoso C, Timsit S, Villard L, Khrestchatisky M, Fontes M, Colleaux L. Specific interaction between the *XNP/ATR-X* gene product and the SET domain of the human EZH2 protein. Hum Mol Genet 1998;7;679-84.

Carlson M, Laurent BC. The SNF/SWI family of global transcriptional activators. Curr Opin Cell Biol 1994;6;396-402.

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 2005;434;400-4.

Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: Exonic mutations that affect splicing. Nat Rev Genet 2002;3;285-98.

Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 2003;31;3568-71.

Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J et al. MeCP2, a key contributor to neurological disease, activates and represses transcription. Science 2008;320;1224-9.

Chahrour M, Zoghbi HY. The story of Rett syndrome: From clinic to neurobiology. Neuron 2007;56;422-37.

Chandler VL, Stam M. Chromatin conversations: Mechanisms and implications of paramutation. Nat Rev Genet 2004;5;532-44.

Chang S, Bray SM, Li Z, Zarnescu DC, He C, Jin P et al. Identification of small molecules rescuing fragile X syndrome phenotypes in Drosophila. Nat Chem Biol 2008;4;256-63.

Chao W, D'Amore PA. IGF2: Epigenetic regulation and role in development and disease. Cytokine Growth Factor Rev 2008;19;111-20.

Chelly J, Khelfaoui M, Francis F, Cherif B, Bienvenu T. Genetics and pathophysiology of mental retardation. Eur J Hum Genet 2006;14;701-13.

Chen W, Jensen LR, Gecz J, Fryns JP, Moraine C, de Brouwer A et al. Mutation screening of brain-expressed X-chromosomal miRNA genes in 464 patients with nonsyndromic X-linked mental retardation. Eur J Hum Genet 2006.

Chen WG, Chang Q, Lin Y, Meissner A, West AE, Griffith EC et al. Derepression of *BDNF* transcription involves calcium-dependent phosphorylation of MeCP2. Science 2003;302;885-9.

Chen Z, Zang J, Whetstine J, Hong X, Davrazou F, Kutateladze TG et al. Structural insights into histone demethylation by JMJD2 family members. Cell 2006;125;691-702.

Chiurazzi P, Schwartz CE, Gecz J, Neri G. XLMR genes: Update 2007. Eur J Hum Genet 2008;16;422-34.

Chiurazzi P, Tabolacci E, Neri G. X-linked mental retardation (XLMR): From clinical conditions to cloned genes. Crit Rev Clin Lab Sci 2004;41;117-58.

Chiyonobu T, Hayashi S, Kobayashi K, Morimoto M, Miyanomae Y, Nishimura A et al. Partial tandem duplication of *GRIA3* in a male with mental retardation. Am J Med Genet A 2007;143A;1448-55.

Cho G, Bhat SS, Gao J, Collins JS, Rogers RC, Simensen RJ et al. Evidence that *SIZN1* is a candidate X-linked mental retardation gene. Am J Med Genet A 2008;146A;2644-50.

Christianson AL, Zwane ME, Manga P, Rosen E, Venter A, Downs D et al. Children with intellectual disability in rural south africa: Prevalence and associated disability. J Intellect Disabil Res 2002;46;179-86.

Claus R, Lubbert M. Epigenetic targets in hematopoietic malignancies. Oncogene 2003;22;6489-96.

Colasante G, Collombat P, Raimondi V, Bonanomi D, Ferrai C, Maira M et al. Arx is a direct target of Dlx2 and thereby contributes to the tangential migration of GABAergic interneurons. J Neurosci 2008;28;10674-86.

Collin RW, de Heer AM, Oostrik J, Pauw RJ, Plantinga RF, Huygen PL et al. Mid-frequency *DFNA8/12* hearing loss caused by a synonymous TECTA mutation that affects an exonic splice enhancer. Eur J Hum Genet 2008;16;1430-6.

Collins FS. Positional cloning moves from perditional to traditional. Nat Genet 1995;9;347-50.

Collombat P, Hecksher-Sorensen J, Broccoli V, Krull J, Ponte I, Mundiger T et al. The simultaneous loss of *Arx* and *Pax4* genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas. Development 2005;132;2969-80.

Colombo E, Collombat P, Colasante G, Bianchi M, Long J, Mansouri A et al. Inactivation of *Arx*, the murine ortholog of the X-linked lissencephaly with ambiguous genitalia gene, leads to severe disorganization of the ventral telencephalon with impaired neuronal migration and differentiation. J Neurosci 2007;27;4786-98.

Colombo E, Galli R, Cossu G, Gecz J, Broccoli V. Mouse orthologue of *ARX*, a gene mutated in several X-linked forms of mental retardation and epilepsy, is a marker of adult neural stem cells and forebrain GABAergic neurons. Dev Dyn 2004;231;631-9.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y et al. Origins and functional impact of copy number variation in the human genome. Nature 2009.

Cooper DN, Krawczak M. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum Genet 1989;83;181-8.

Cornish K, Turk J, Hagerman R. The fragile X continuum: New advances and perspectives. J Intellect Disabil Res 2008;52;469-82.

Costello JF, Plass C. Methylation matters. J Med Genet 2001;38;285-303.

Couvert P, Bienvenu T, Aquaviva C, Poirier K, Moraine C, Gendrot C et al. *MECP2* is highly mutated in X-linked mental retardation. Hum Mol Genet 2001;10;941-6.

de Brouwer AP, Yntema HG, Kleefstra T, Lugtenberg D, Oudakker AR, de Vries BB et al. Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. Hum Mutat 2007;28;207-8.

De La Fuente R, Viveiros MM, Wigglesworth K, Eppig JJ. ATRX, a member of the SNF2 family of helicase/ATPases, is required for chromosome alignment and meiotic spindle organization in metaphase II stage mouse oocytes. Dev Biol 2004;272;1-14.

de Vries BB, Pfundt R, Leisink M, Koolen DA, Vissers LE, Janssen IM et al. Diagnostic genome profiling in mental retardation. Am J Hum Genet 2005;77;606-16.

Demirci FY, White NJ, Rigatti BW, Lewis KF, Gorin MB. Identification, genomic structure, and screening of the vacuolar proton-ATPase membrane sector-associated protein M8-9 gene within the COD1 critical region (Xp11.4). Mol Vis 2001;7;234-9.

Demos MK, Fullston T, Partington MW, Gecz J, Gibson WT. Clinical study of two brothers with a novel 33 bp duplication in the *ARX* gene. Am J Med Genet A 2009;149A;1482-6.

Deplus R, Brenner C, Burgers WA, Putmans P, Kouzarides T, de Launoit Y et al. DNMT3L is a transcriptional repressor that recruits histone deacetylase. Nucleic Acids Res 2002;30;3831-8.

D'Hulst C, Kooy RF. Fragile X syndrome: From molecular genetics to therapy. J Med Genet 2009;46;577-84.

Dick KJ, Eckhardt M, Paisan-Ruiz C, Alshehhi AA, Proukakis C, Sibtain NA et al. Mutation of FA2H underlies a complicated form of hereditary spastic paraplegia (SPG35). Hum Mutat 2010;31;E1251-60.

Dimos JT, Rodolfa KT, Niakan KK, Weisenthal LM, Mitsumoto H, Chung W et al. Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science 2008;321;1218-21.

Durkin M. The epidemiology of developmental disabilities in low-income countries. Ment Retard Dev Disabil Res Rev 2002;8;206-11.

Ehrlich M. The ICF syndrome, a DNA methyltransferase 3B deficiency and immunodeficiency disease. Clin Immunol 2003;109;17-28.

Ehrlich M, Sanchez C, Shao C, Nishiyama R, Kehrl J, Kuick R et al. ICF, an immunodeficiency syndrome: DNA methyltransferase 3B involvement, chromosome anomalies, and gene dysregulation. Autoimmunity 2008;41;253-71.

ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. Science 2004;306;636-40.

ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007;447;799-816.

Engels H, Brockschmidt A, Hoischen A, Landwehr C, Bosse K, Walldorf C et al. DNA microarray analysis identifies candidate regions and genes in unexplained mental retardation. Neurology 2007;68;743-50.

Fan YS, Jayakar P, Zhu H, Barbouth D, Sacharow S, Morales A et al. Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization. Hum Mutat 2007;28;1124-32.

Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to crohn disease of the colon. Am J Hum Genet 2006;79;439-48.

Field M, Tarpey PS, Smith R, Edkins S, O'Meara S, Stevens C et al. Mutations in the *BRWD3* gene cause X-linked mental retardation associated with macrocephaly. Am J Hum Genet 2007;81;367-74.

Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D et al. DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet 2009;84;524-33.

Flint J, Knight S. The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation. Curr Opin Genet Dev 2003;13;310-6.

Friedman J, Adam S, Arbour L, Armstrong L, Baross A, Birch P et al. Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500 K SNP array genomic hybridization. BMC Genomics 2009;10;526.

Friedman JM, Baross A, Delaney AD, Ally A, Arbour L, Armstrong L et al. Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. Am J Hum Genet 2006;79;500-13.

Frints SG, Froyen G, Marynen P, Willekens D, Legius E, Fryns JP. Re-evaluation of MRX36 family after discovery of an *ARX* gene mutation reveals mild neurological features of partington syndrome. Am J Med Genet 2002;112;427-8.

Froyen G, Bauters M, Voet T, Marynen P. X-linked mental retardation and epigenetics. J Cell Mol Med 2006;10;808-25.

Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H et al. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein *MBD1* isoforms. Mol Cell Biol 1999;19;6415-26.

Fullenkamp AN, El-Hodiri HM. The function of the aristaless-related homeobox (*Arx*) gene product as a transcriptional repressor is diminished by mutations associated with X-linked mental retardation (XLMR). Biochem Biophys Res Commun 2008;377;73-8.

Fullston T, Brueton L, Willis T, Philip S, MacPherson L, Finnis M et al. Ohtahara syndrome in a family with an ARX protein truncation mutation (c.81C>G/p.Y27X). Eur J Hum Genet 2010;18;157-62.

Fulp CT, Cho G, Marsh ED, Nasrallah IM, Labosky PA, Golden JA. Identification of Arx transcriptional targets in the developing basal forebrain. Hum Mol Genet 2008;17;3740-60.

Garrick D, Samara V, McDowell TL, Smith AJ, Dobbie L, Higgs DR et al. A conserved truncated isoform of the ATR-X syndrome protein lacking the SWI/SNF-homology domain. Gene 2004;326;23-34.

Garrick D, Sharpe JA, Arkell R, Dobbie L, Smith AJ, Wood WG et al. Loss of Atrx affects trophoblast development and the pattern of X-inactivation in extraembryonic tissues. PLoS Genet 2006;2;e58.

Gecz J, Cloosterman D, Partington M. *ARX*: A gene for all seasons. Curr Opin Genet Dev 2006;16;308-16.

Gecz J, Shoubridge C, Corbett M. The genetic landscape of intellectual disability arising from chromosome X. Trends Genet 2009;25;308-16.

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S et al. Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 2004;5;R80.

Gestinari-Duarte Rde S, Santos-Reboucas CB, Boy RT, Pimentel MM. *ARX* mutation c.428-451dup (24bp) in a brazilian family with X-linked mental retardation. Eur J Med Genet 2006;49;269-75.

Giacometti E, Luikenhuis S, Beard C, Jaenisch R. Partial rescue of MeCP2 deficiency by postnatal activation of MeCP2. Proc Natl Acad Sci U S A 2007;104;1931-6.

Gibbons R. Alpha thalassaemia-mental retardation, X linked. Orphanet J Rare Dis 2006;1;15.

Gibbons RJ, Higgs DR. Molecular-clinical spectrum of the ATR-X syndrome. Am J Med Genet 2000;97;204-12.

Gibbons RJ, McDowell TL, Raman S, O'Rourke DM, Garrick D, Ayyub H et al. Mutations in *ATRX*, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation. Nat Genet 2000;24;368-71.

Gibbons RJ, Picketts DJ, Villard L, Higgs DR. Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome). Cell 1995;80;837-45.

Gibbons RJ, Wada T, Fisher CA, Malik N, Mitson MJ, Steensma DP et al. Mutations in the chromatin-associated protein *ATRX*. Hum Mutat 2008;29;796-802.

Goldstone AP. Prader-willi syndrome: Advances in genetics, pathophysiology and treatment. Trends Endocrinol Metab 2004;15;12-20.

Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X et al. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. Science 2006;311;395-8.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G et al. The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 2005;307;1434-40.

Gronskov K, Hjalgrim H, Nielsen IM, Brondum-Nielsen K. Screening of the *ARX* gene in 682 retarded males. Eur J Hum Genet 2004;12;701-5.

Guerrini R, Moro F, Kato M, Barkovich AJ, Shiihara T, McShane MA et al. Expansion of the first PolyA tract of ARX causes infantile spasms and status dystonicus. Neurology 2007;69;427-33.

Guerrini R, Shanahan JL, Carrozzo R, Bonanni P, Higgs DR, Gibbons RJ. A nonsense mutation of the *ATRX* gene causing mild mental retardation and epilepsy. Ann Neurol 2000;47;117-21.

Guy J, Gan J, Selfridge J, Cobb S, Bird A. Reversal of neurological defects in a mouse model of rett syndrome. Science 2007;315;1143-7.

Hagerman RJ, Berry-Kravis E, Kaufmann WE, Ono MY, Tartaglia N, Lachiewicz A et al. Advances in the treatment of fragile X syndrome. Pediatrics 2009;123;378-90.

Harum KH, Alemi L, Johnston MV. Cognitive impairment in Coffin-Lowry syndrome correlates with reduced *RSK2* activation. Neurology 2001;56;207-14.

Hedera P, Alvarado D, Beydoun A, Fink JK. Novel mental retardation-epilepsy syndrome linked to Xp21.1-p11.4. Ann Neurol 2002;51;45-50.

Hehir-Kwa JY, Egmont-Petersen M, Janssen IM, Smeets D, van Kessel AG, Veltman JA. Genome-wide copy number profiling on high-density bacterial artificial chromosomes, single-nucleotide polymorphisms, and oligonucleotide microarrays: A platform comparison based on statistical power analysis. DNA Res 2007;14;1-11.

Hellman A, Chess A. Gene body-specific methylation on the active X chromosome. Science 2007;315;1141-3.

Hemler ME. Tetraspanin functions and associated microdomains. Nat Rev Mol Cell Biol 2005;6;801-11.

Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 1998;18;6538-47.

Hendriks RW, Chen ZY, Hinds H, Schuurman RK, Craig IW. An X chromosome inactivation assay based on differential methylation of a CpG island coupled to a VNTR polymorphism at the 5' end of the monoamine oxidase A gene. Hum Mol Genet 1992;1;662.

Higgins JJ, Pucilowska J, Lombardi RQ, Rooney JP. A mutation in a novel ATP-dependent lon protease gene in a kindred with mild mental retardation. Neurology 2004;63;1927-31.

Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 2007;356;697-705.

Holliday R. The inheritance of epigenetic defects. Science 1987;238;163-70.

Holliday R. Epigenetics: A historical overview. Epigenetics 2006;1;76-80.

Hosoki K, Ogata T, Kagami M, Tanaka T, Saitoh S. Epimutation (hypomethylation) affecting the chromosome 14q32.2 imprinted region in a girl with upd(14)mat-like phenotype. Eur J Hum Genet 2008;16;1019-23.

Host L, Dietrich JB, Carouge D, Aunis D, Zwiller J. Cocaine self-administration alters the expression of chromatin-remodelling proteins; modulation by histone deacetylase inhibition. J Psychopharmacol 2009.

Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S et al. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat 2008;29;975-82.

Hoyer J, Dreweke A, Becker C, Gohring I, Thiel CT, Peippo MM et al. Molecular karyotyping in patients with mental retardation using 100K single-nucleotide polymorphism arrays. J Med Genet 2007;44;629-36.

Ichimura T, Watanabe S, Sakamoto Y, Aoto T, Fujita N, Nakao M. Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. J Biol Chem 2005;280;13928-35.

Inlow JK, Restifo LL. Molecular and comparative genetics of mental retardation. Genetics 2004;166;835-81.

Ishov AM, Vladimirova OV, Maul GG. Heterochromatin and ND10 are cell-cycle regulated and phosphorylation-dependent alternate nuclear sites of the transcription repressor Daxx and SWI/SNF protein ATRX. J Cell Sci 2004;117;3807-20.

Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH et al. The X-linked mental retardation gene *SMCX*/*JARID1C* defines a family of histone H3 lysine 4 demethylases. Cell 2007;128;1077-88.

Jaillard S, Drunat S, Bendavid C, Aboura A, Etcheverry A, Journel H et al. Identification of gene copy number variations in patients with mental retardation using array-CGH: Novel syndromes in a large french series. Eur J Med Genet 2009.

Jensen LR, Amende M, Gurok U, Moser B, Gimmel V, Tzschach A et al. Mutations in the JARID1C gene, which is involved in transcriptional regulation and chromatin remodeling, cause X-linked mental retardation. Am J Hum Genet 2005;76;227-36.

Jones PA, Baylin SB. The fundamental role of epigenetic events in cancer. Nat Rev Genet 2002;3;415-28.

Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nat Genet 1998;19;187-91.

Jugloff DG, Vandamme K, Logan R, Visanji NP, Brotchie JM, Eubanks JH. Targeted delivery of an Mecp2 transgene to forebrain neurons improves the behavior of female Mecp2-deficient mice. Hum Mol Genet 2008;17;1386-96.

Kafri T, Ariel M, Brandeis M, Shemer R, Urven L, McCarrey J et al. Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. Genes Dev 1992;6;705-14.

Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H et al. Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. Cancer Res 1997;57;808-11.

Kangaspeska S, Stride B, Metivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP et al. Transient cyclical methylation of promoter DNA. Nature 2008;452;112-5.

Kato M, Das S, Petras K, Kitamura K, Morohashi K, Abuelo DN et al. Mutations of *ARX* are associated with striking pleiotropy and consistent genotype-phenotype correlation. Hum Mutat 2004;23;147-59.

Kato M, Das S, Petras K, Sawaishi Y, Dobyns WB. Polyalanine expansion of ARX associated with cryptogenic West syndrome. Neurology 2003;61;267-76.

Kato M, Saitoh S, Kamei A, Shiraishi H, Ueda Y, Akasaka M et al. A longer polyalanine expansion mutation in the *ARX* gene causes early infantile epileptic encephalopathy with suppression-burst pattern (Ohtahara syndrome). Am J Hum Genet 2007;81;361-6.

Khalil AM, Faghihi MA, Modarresi F, Brothers SP, Wahlestedt C. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. PLoS One 2008;3;e1486.

Kim MS, Kondo T, Takada I, Youn MY, Yamamoto Y, Takahashi S et al. DNA demethylation in hormone-induced transcriptional derepression. Nature 2009;461;1007-12.

King RA, Rotter JI, Motulsky AG. The genetic basis of common diseases. 2nd edition. Oxford: Oxford University Press; 2002.

Kitamura K, Itou Y, Yanazawa M, Ohsawa M, Suzuki-Migishima R, Umeki Y et al. Three human *ARX* mutations cause the lissencephaly-like and mental retardation with epilepsy-like pleiotropic phenotypes in mice. Hum Mol Genet 2009;18;3708-24.

Kitamura K, Yanazawa M, Sugiyama N, Miura H, Iizuka-Kogo A, Kusaka M et al. Mutation of *ARX* causes abnormal development of forebrain and testes in mice and X-linked lissencephaly with abnormal genitalia in humans. Nat Genet 2002;32;359-69.

Kleefstra T, Hamel BC. X-linked mental retardation: Further lumping, splitting and emerging phenotypes. Clin Genet 2005;67;451-67.

Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S et al. Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 1998;4;844-7.

Koolen DA, Sharp AJ, Hurst JA, Firth HV, Knight SJ, Goldenberg A et al. Clinical and molecular delineation of the 17q21.31 microdeletion syndrome. J Med Genet 2008;45;710-20.

Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 2006;38;999-1001.

Kortschak RD, Tucker PW, Saint R. ARID proteins come in from the desert. Trends Biochem Sci 2000;25;294-9.

Kousoulidou L, Parkel S, Zilina O, Palta P, Puusepp H, Remm M et al. Screening of 20 patients with X-linked mental retardation using chromosome X-specific array-MAPH. Eur J Med Genet 2007;50;399-410.

Kouzarides T. Chromatin modifications and their function. Cell 2007;128;693-705.

Kramer JM, van Bokhoven H. Genetic and epigenetic defects in mental retardation. Int J Biochem Cell Biol 2009;41;96-107.

Krepischi-Santos AC, Vianna-Morgante AM, Jehee FS, Passos-Bueno MR, Knijnenburg J, Szuhai K et al. Whole-genome array-CGH screening in undiagnosed syndromic patients: Old syndromes revisited and new alterations. Cytogenet Genome Res 2006;115;254-61.

Kuhn DE, Nuovo GJ, Martin MM, Malana GE, Pleister AP, Jiang J et al. Human chromosome 21-derived miRNAs are overexpressed in down syndrome brains and hearts. Biochem Biophys Res Commun 2008;370;473-7.

Kuhn DE, Nuovo GJ, Terry AV,Jr, Martin MM, Malana GE, Sansom SE et al. Chromosome 21-derived microRNAs provide an etiological basis for aberrant protein expression in human down syndrome brains. J Biol Chem 2010;285;1529-43.

Kuwabara T, Hsieh J, Nakashima K, Taira K, Gage FH. A small modulatory dsRNA specifies the fate of adult neural stem cells. Cell 2004;116;779-93.

Ladd PD, Smith LE, Rabaia NA, Moore JM, Georges SA, Hansen RS et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum Mol Genet 2007;16;3174-87.

Laperuta C, Spizzichino L, D'Adamo P, Monfregola J, Maiorino A, D'Eustacchio A et al. MRX87 family with aristaless X dup24bp mutation and implication for polyAlanine expansions. BMC Med Genet 2007;8;25.

LaSalle JM, Goldstine J, Balmer D, Greco CM. Quantitative localization of heterogeneous methyl-CpG-binding protein 2 (MeCP2) expression phenotypes in normal and rett syndrome brain by laser scanning cytometry. Hum Mol Genet 2001;10;1729-40.

Laumonnier F, Ronce N, Hamel BC, Thomas P, Lespinasse J, Raynaud M et al. Transcription factor *SOX3* is involved in X-linked mental retardation with growth hormone deficiency. Am J Hum Genet 2002;71;1450-5.

Lavoie H, Debeane F, Trinh QD, Turcotte JF, Corbeil-Girard LP, Dicaire MJ et al. Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domains. Hum Mol Genet 2003;12;2967-79.

Leonard H, Wen X. The epidemiology of mental retardation: Challenges and opportunities in the new millennium. Ment Retard Dev Disabil Res Rev 2002;8;117-34.

Levenson JM, O'Riordan KJ, Brown KD, Trinh MA, Molfese DL, Sweatt JD. Regulation of histone acetylation during memory formation in the hippocampus. J Biol Chem 2004;279;40545-59.

Levenson JM, Roth TL, Lubin FD, Miller CA, Huang IC, Desai P et al. Evidence that DNA (cytosine-5) methyltransferase regulates synaptic plasticity in the hippocampus. J Biol Chem 2006;281;15763-73.

Levenson JM, Sweatt JD. Epigenetic mechanisms in memory formation. Nat Rev Neurosci 2005;6;108-18.

Levy MA, Fernandes AD, Tremblay DC, Seah C, Berube NG. The SWI/SNF protein *ATRX* co-regulates pseudoautosomal genes that have translocated to autosomes in the mouse genome. BMC Genomics 2008;9;468.

Li N, Carrel L. Escape from X chromosome inactivation is an intrinsic property of the Jarid1c locus. Proc Natl Acad Sci U S A 2008;105;17055-60.

Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M et al. Heritable somatic methylation and inactivation of *MSH2* in families with lynch syndrome due to deletion of the 3' exons of *TACSTD1*. Nat Genet 2009;41;112-7.

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-delta delta C(T)) method. Methods 2001;25;402-8.

Lossi AM, Millan JM, Villard L, Orellana C, Cardoso C, Prieto F et al. Mutation of the *XNP*/*ATR-X* gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: Demonstration that the mutation is involved in the inactivation bias. Am J Hum Genet 1999;65;558-62.

Lossie AC, Whitney MM, Amidon D, Dong HJ, Chen P, Theriaque D et al. Distinct phenotypes distinguish the molecular classes of Angelman syndrome. J Med Genet 2001;38;834-45.

Louie LG, King MC. A novel approach to establishing permanent lymphoblastoid cell lines: Epstein-barr virus transformation of cryopreserved lymphocytes. Am J Hum Genet 1991;48;637-8.

Lu X, Shaw CA, Patel A, Li J, Cooper ML, Wells WR et al. Clinical implementation of chromosomal microarray analysis: Summary of 2513 postnatal cases. PLoS One 2007;2;e327.

Lugtenberg D, de Brouwer AP, Kleefstra T, Oudakker AR, Frints SG, Schrander-Stumpel CT et al. Chromosomal copy number changes in patients with non-syndromic X linked mental retardation detected by array CGH. J Med Genet 2006;43;362-70.

Lugtenberg D, de Brouwer AP, Oudakker AR, Pfundt R, Hamel BC, van Bokhoven H et al. Xq13.2q21.1 duplication encompassing the *ATRX* gene in a man with mental retardation, minor facial and genital anomalies, short stature and broad thorax. Am J Med Genet A 2009;149A;760-6.

Madrigal I, Rodriguez-Revenga L, Armengol L, Gonzalez E, Rodriguez B, Badenas C et al. X-chromosome tiling path array detection of copy number variants in patients with chromosome X-linked mental retardation. BMC Genomics 2007;8;443.

Maher ER, Reik W. Beckwith-Wiedemann syndrome: Imprinting in clusters revisited. J Clin Invest 2000;105;247-52.

Makedonski K, Abuhatzira L, Kaufman Y, Razin A, Shemer R. MeCP2 deficiency in rett syndrome causes epigenetic aberrations at the PWS/AS imprinting center that affects *UBE3A* expression. Hum Mol Genet 2005;14;1049-58.

Mamedov TG, Pienaar E, Whitney SE, TerMaat JR, Carvill G, Goliath R et al. A fundamental study of the PCR amplification of GC-rich DNA templates. Comput Biol Chem 2008;32;452-7.

Mandel JL, Chelly J. Monogenic X-linked mental retardation: Is it as frequent as currently estimated? The paradox of the *ARX* (aristaless X) mutations. Eur J Hum Genet 2004;12;689-93.

Martinez-Garay I, Tomas M, Oltra S, Ramser J, Molto MD, Prieto F et al. A two base pair deletion in the *PQBP1* gene is associated with microphthalmia, microcephaly, and mental retardation. Eur J Hum Genet 2007;15;29-34.

Martinowich K, Hattori D, Wu H, Fouse S, He F, Hu Y et al. DNA methylation-related chromatin remodeling in activity-dependent *BDNF* gene regulation. Science 2003;302;890-3.

May PA, Gossage JP, Marais AS, Adnams CM, Hoyme HE, Jones KL et al. The epidemiology of fetal alcohol syndrome and partial FAS in a South African community. Drug Alcohol Depend 2007;88;259-71.

McBride SM, Choi CH, Wang Y, Liebelt D, Braunstein E, Ferreiro D et al. Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a drosophila model of fragile X syndrome. Neuron 2005;45;753-64.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC et al. Common deletion polymorphisms in the human genome. Nat Genet 2006;38;86-92.

McDowell TL, Gibbons RJ, Sutherland H, O'Rourke DM, Bickmore WA, Pombo A et al. Localization of a putative transcriptional regulator (*ATRX*) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. Proc Natl Acad Sci U S A 1999;96;13983-8.

McKenzie O, Ponte I, Mangelsdorf M, Finnis M, Colasante G, Shoubridge C et al. Aristaless-related homeobox gene, the gene responsible for west syndrome and related

disorders, is a Groucho/transducin-like enhancer of split dependent transcriptional repressor. Neuroscience 2007;146;236-47.

McKinley MJ, Albiston AL, Allen AM, Mathai ML, May CN, McAllen RM et al. The brain renin-angiotensin system: Location and physiological roles. Int J Biochem Cell Biol 2003;35;901-18.

McLaren J, Bryson SE. Review of recent epidemiological studies of mental retardation: Prevalence, associated disorders, and etiology. Am J Ment Retard 1987;92;243-54.

McMullan DJ, Bonin M, Hehir-Kwa JY, de Vries BB, Dufke A, Rattenberry E et al. Molecular karyotyping of patients with unexplained mental retardation by SNP arrays: A multicenter study. Hum Mutat 2009;30;1082-92.

Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. N Engl J Med 2008;359;1685-99.

Menten B, Maas N, Thienpont B, Buysse K, Vandesompele J, Melotte C et al. Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: A new series of 140 patients and review of published reports. J Med Genet 2006;43;625-33.

Metivier R, Gallais R, Tiffoche C, Le Peron C, Jurkowska RZ, Carmouche RP et al. Cyclical DNA methylation of a transcriptionally active promoter. Nature 2008;452;45-50.

Miller CA, Campbell SL, Sweatt JD. DNA methylation and histone acetylation work in concert to regulate memory formation and synaptic plasticity. Neurobiol Learn Mem 2008;89;599-603.

Miller CA, Sweatt JD. Covalent modification of DNA regulates memory formation. Neuron 2007;53;857-69.

Miller DT, Shen Y, Weiss LA, Korn J, Anselm I, Bridgemohan C et al. Microdeletion/duplication at 15q13.2q13.3 among individuals with features of autism and other neuropsychiatric disorders. J Med Genet 2009;46;242-8.

Mir A, Kaufman L, Noor A, Motazacker MM, Jamil T, Azam M et al. Identification of mutations in *TRAPPC9*, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am J Hum Genet 2009;85;909-15.

Miyake N, Shimokawa O, Harada N, Sosonkina N, Okubo A, Kawara H et al. BAC array CGH reveals genomic aberrations in idiopathic mental retardation. Am J Med Genet A 2006;140;205-11.

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 2008;30;755-66.

Molinari F, Foulquier F, Tarpey PS, Morelle W, Boissel S, Teague J et al. Oligosaccharyltransferase-subunit mutations in nonsyndromic mental retardation. Am J Hum Genet 2008;82;1150-7.

Molinari F, Rio M, Meskenaite V, Encha-Razavi F, Auge J, Bacq D et al. Truncating neurotrypsin mutation in autosomal recessive nonsyndromic mental retardation. Science 2002;298;1779-81.

Monk M, Boubelik M, Lehnert S. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development 1987;99;371-82.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E. Epigenetic inheritance at the agouti locus in the mouse. Nat Genet 1999;23;314-8.

Motazacker MM, Rost BR, Hucho T, Garshasbi M, Kahrizi K, Ullmann R et al. A defect in the ionotropic glutamate receptor 6 gene (*GRIK2*) is associated with autosomal recessive mental retardation. Am J Hum Genet 2007;81;792-8.

Mulley JC, Kerr B, Stevenson R, Lubs H. Nomenclature guidelines for X-linked mental retardation. Am J Med Genet 1992;43;383-91.

Murakami K, Ohhira T, Oshiro E, Qi D, Oshimura M, Kugoh H. Identification of the chromatin regions coated by non-coding xist RNA. Cytogenet Genome Res 2009;125;19-25.

Musante L, Kunde SA, Sulistio TO, Fischer U, Grimme A, Frints SG et al. Common pathological mutations in *PQBP1* induce nonsense-mediated mRNA decay and enhance exclusion of the mutant exon. Hum Mutat 2010;31;90-8.

Mutskov V, Felsenfeld G. Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO J 2004;23;138-49.

Nan X, Hou J, Maclean A, Nasir J, Lafuente MJ, Shu X et al. Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. Proc Natl Acad Sci U S A 2007;104;2709-14.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 1998;393;386-9.

Nascimento RM, Otto PA, de Brouwer AP, Vianna-Morgante AM. *UBE2A*, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. Am J Hum Genet 2006;79;549-55.

Nasrallah IM, Minarcik JC, Golden JA. A polyalanine tract expansion in ARX forms intranuclear inclusions and results in increased cell death. J Cell Biol 2004;167;411-6.

Nawara M, Szczaluba K, Poirier K, Chrzanowska K, Pilch J, Bal J et al. The *ARX* mutations: A frequent cause of X-linked mental retardation. Am J Med Genet A 2006;140;727-32.

Nguyen G, Delarue F, Burckle C, Bouzhir L, Giller T, Sraer JD. Pivotal role of the renin/prorenin receptor in angiotensin II production and cellular responses to renin. J Clin Invest 2002;109;1417-27.

Nishiyama A, Takeshima Y, Zhang Z, Habara Y, Tran TH, Yagi M et al. Dystrophin nonsense mutations can generate alternative rescue transcripts in lymphocytes. Ann Hum Genet 2008;72;717-24.

Nomura T, Kimura M, Horii T, Morita S, Soejima H, Kudo S et al. MeCP2-dependent repression of an imprinted miR-184 released by depolarization. Hum Mol Genet 2008;17;1192-9.

Nowakowska B, Stankiewicz P, Obersztyn E, Ou Z, Li J, Chinault AC et al. Application of metaphase HR-CGH and targeted chromosomal microarray analyses to genomic characterization of 116 patients with mental retardation and dysmorphic features. Am J Med Genet A 2008;146A;2361-9.

Ohira R, Zhang YH, Guo W, Dipple K, Shih SL, Doerr J et al. Human *ARX* gene: Genomic characterization and expression. Mol Genet Metab 2002;77;179-88.

Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 1999;99;247-57.

Okano M, Xie S, Li E. Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucleic Acids Res 1998;26;2536-40.

Oti M, Brunner HG. The modular nature of genetic diseases. Clin Genet 2007;71;1-11.

Park DJ, Pask AJ, Huynh K, Renfree MB, Harley VR, Graves JA. Comparative analysis of *ATRX*, a chromatin remodeling protein. Gene 2004;339;39-48.

Partington MW, Turner G, Boyle J, Gecz J. Three new families with X-linked mental retardation caused by the 428-451dup(24bp) mutation in *ARX*. Clin Genet 2004;66;39-45.

Parvari R, Mumm S, Galil A, Manor E, Bar-David Y, Carmi R. Deletion of 8.5 mb, including the *FMR1* gene, in a male with the fragile X syndrome phenotype and overgrowth. Am J Med Genet 1999;83;302-7.

Pelizzola M, Koga Y, Urban AE, Krauthammer M, Weissman S, Halaban R et al. MEDME: An experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. Genome Res 2008;18;1652-9.

Petrij F, Giles RH, Dauwerse HG, Saris JJ, Hennekam RC, Masuno M et al. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. Nature 1995;376;348-51.

Picketts DJ, Higgs DR, Bachoo S, Blake DJ, Quarrell OW, Gibbons RJ. *ATRX* encodes a novel member of the SNF2 family of proteins: Mutations point to a common mechanism underlying the ATR-X syndrome. Hum Mol Genet 1996;5;1899-907.

Pietrobono R, Pomponi MG, Tabolacci E, Oostra B, Chiurazzi P, Neri G. Quantitative analysis of DNA demethylation and transcriptional reactivation of the *FMR1* gene in fragile X cells treated with 5-azadeoxycytidine. Nucleic Acids Res 2002;30;3278-85.

Pietrobono R, Tabolacci E, Zalfa F, Zito I, Terracciano A, Moscato U et al. Molecular dissection of the events leading to inactivation of the *FMR1* gene. Hum Mol Genet 2005;14;267-77.

Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. Am J Hum Genet 2002;71;168-73.

Plenge RM, Tranebjaerg L, Jensen PK, Schwartz C, Willard HF. Evidence that mutations in the X-linked DDP gene cause incompletely penetrant and variable skewed X inactivation. Am J Hum Genet 1999;64;759-67.

Poirier K, Abriol J, Souville I, Laroche-Raynaud C, Beldjord C, Gilbert B et al. Maternal mosaicism for mutations in the *ARX* gene in a family with X linked mental retardation. Hum Genet 2005;118;45-8.

Poirier K, Eisermann M, Caubel I, Kaminska A, Peudonnier S, Boddaert N et al. Combination of infantile spasms, non-epileptic seizures and complex movement disorder: A new case of *ARX*-related epilepsy. Epilepsy Res 2008;80;224-8.

Poirier K, Lacombe D, Gilbert-Dussardier B, Raynaud M, Desportes V, de Brouwer AP et al. Screening of *ARX* in mental retardation families: Consequences for the strategy of molecular diagnosis. Neurogenetics 2006;7;39-46.

Poirier K, Van Esch H, Friocourt G, Saillour Y, Bahi N, Backer S et al. Neuroanatomical distribution of *ARX* in brain and its localisation in GABAergic neurons. Brain Res Mol Brain Res 2004;122;35-46.

Poot M, Eleveld MJ, van 't Slot R, Ploos van Amstel HK, Hochstenbach R. Recurrent copy number changes in mentally retarded children harbour genes involved in cellular localization and the glutamate receptor complex. Eur J Hum Genet 2009.

Ramser J, Abidi FE, Burckle CA, Lenski C, Toriello H, Wen G et al. A unique exonic splice enhancer mutation in a family with X-linked mental retardation and epilepsy points to a novel role of the renin receptor. Hum Mol Genet 2005;14;1019-27.

Rastegar M, Hotta A, Pasceri P, Makarem M, Cheung AY, Elliott S et al. *MECP2* isoform-specific vectors with regulated expression for rett syndrome gene therapy. PLoS One 2009;4;e6810.

Raymond FL. X linked mental retardation: A clinical guide. J Med Genet 2006;43;193-200.

Raymond FL, Tarpey P. The genetics of mental retardation. Hum Mol Genet 2006;15 Spec No 2;R110-6.

Razin A, Riggs AD. DNA methylation and gene function. Science 1980;210;604-10.

Razin A, Szyf M. DNA methylation patterns. formation and function. Biochim Biophys Acta 1984;782;331-42.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD et al. Global variation in copy number in the human genome. Nature 2006;444;444-54.

Reish O, Fullston T, Regev M, Heyman E, Gecz J. A novel de novo 27 bp duplication of the *ARX* gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations. Am J Med Genet A 2009;149A;1655-60.

Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene 2009;28;1653-68.

Ridout CK, Keighley P, Krywawych S, Brown RM, Brown GK. A putative exonic splicing enhancer in exon 7 of the *PDHA1* gene affects splicing of adjacent exons. Hum Mutat 2008;29;451.

Ritchie K, Seah C, Moulin J, Isaac C, Dick F, Berube NG. Loss of ATRX leads to chromosome cohesion and congression defects. J Cell Biol 2008;180;315-24.

Rizzoti K, Brunelli S, Carmignac D, Thomas PQ, Robinson IC, Lovell-Badge R. SOX3 is required during the formation of the hypothalamo-pituitary axis. Nat Genet 2004;36;247-55.

Roberts LJ, Ramesar RS, Greenberg J. Clinical utility of the ABCR400 microarray: Basing a genetic service on a commercial gene chip. Arch Ophthalmol 2009;127;549-54.

Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with RB, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 2000;25;338-42.

Roelfsema JH, White SJ, Ariyurek Y, Bartholdi D, Niedrist D, Papadia F et al. Genetic heterogeneity in rubinstein-taybi syndrome: Mutations in both the *CBP* and *EP300* genes cause disease. Am J Hum Genet 2005;76;572-80.

Rollins RA, Haghighi F, Edwards JR, Das R, Zhang MQ, Ju J et al. Large-scale structure of genomic methylation patterns. Genome Res 2006;16;157-63.

Ropers HH. X-linked mental retardation: Many genes for a complex disorder. Curr Opin Genet Dev 2006;16;260-9.

Ropers HH, Hamel BC. X-linked mental retardation. Nat Rev Genet 2005;6;46-57.

Rosenberg C, Knijnenburg J, Bakker E, Vianna-Morgante AM, Sloos W, Otto PA et al. Array-CGH detection of micro rearrangements in mentally retarded individuals: Clinical significance of imbalances present both in affected children and normal parents. J Med Genet 2006;43;180-6.

Rosenberg C, Knijnenburg J, Chauffaille Mde L, Brunoni D, Catelani AL, Sloos W et al. Array CGH detection of a cryptic deletion in a complex chromosome rearrangement. Hum Genet 2005;116;390-4.

Rousseau F, Vincent A, Rivella S, Heitz D, Triboli C, Maestrini E et al. Four chromosomal breakpoints and four new probes mark out a 10-cM region encompassing the fragile-X locus (FRAXA). Am J Hum Genet 1991;48;108-16.

Rujirabanjerd S, Nelson J, Tarpey PS, Hackett A, Edkins S, Raymond FL et al. Identification and characterization of two novel *JARID1C* mutations: Suggestion of an emerging genotype-phenotype correlation. Eur J Hum Genet 2009.

Rujirabanjerd S, Tongsippunyoo K, Sripo T, Limprasert P. Mutation screening of the aristaless-related homeobox (*ARX*) gene in thai pediatric patients with delayed development: First report from thailand. Eur J Med Genet 2007;50;346-54.

Santos C, Rodriguez-Revenga L, Madrigal I, Badenas C, Pineda M, Mila M. A novel mutation in *JARID1C* gene associated with mental retardation. Eur J Hum Genet 2006;14;583-6.

Scheffer IE, Wallace RH, Phillips FL, Hewson P, Reardon K, Parasivam G et al. X-linked myoclonic epilepsy with spasticity and intellectual disability: Mutation in the homeobox gene *ARX*. Neurology 2002;59;348-56.

Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M. Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). J Med Genet 2005;42;699-705.

Schuettengruber B, Chourrout D, Vervoort M, Leblanc B, Cavalli G. Genome regulation by polycomb and trithorax proteins. Cell 2007;128;735-45.

Seufert DW, Prescott NL, El-Hodiri HM. Xenopus aristaless-related homeobox (xARX) gene product functions as both a transcriptional activator and repressor in forebrain development. Dev Dyn 2005;232;313-24.

Shaffer LG, Bejjani BA, Torchia B, Kirkpatrick S, Coppinger J, Ballif BC. The identification of microdeletion syndromes and other chromosome abnormalities: Cytogenetic methods of the past, new technologies for the future. Am J Med Genet C Semin Med Genet 2007;145C;335-45.

Shaffer LG, Kashork CD, Saleki R, Rorem E, Sundin K, Ballif BC et al. Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. J Pediatr 2006;149;98-102.

Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis 2010;31;27-36.

Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. Annu Rev Genomics Hum Genet 2006;7;407-42.

Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA et al. Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 2006;38;1038-42.

Sharp AJ, Selzer RR, Veltman JA, Gimelli S, Gimelli G, Striano P et al. Characterization of a recurrent 15q24 microdeletion syndrome. Hum Mol Genet 2007;16;567-72.

Shaw-Smith C, Redon R, Rickman L, Rio M, Willatt L, Fiegler H et al. Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. J Med Genet 2004;41;241-8.

Sheardown S, Norris D, Fisher A, Brockdorff N. The mouse smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. Hum Mol Genet 1996;5;1355-60.

Shen Y, Irons M, Miller DT, Cheung SW, Lip V, Sheng X et al. Development of a focused oligonucleotide-array comparative genomic hybridization chip for clinical diagnosis of genomic imbalance. Clin Chem 2007;53;2051-9.

Shinozaki Y, Osawa M, Sakuma H, Komaki H, Nakagawa E, Sugai K et al. Expansion of the first polyalanine tract of the *ARX* gene in a boy presenting with generalized dystonia in the absence of infantile spasms. Brain Dev 2009;31;469-72.

Shoubridge C, Cloosterman D, Parkinson-Lawerence E, Brooks D, Gecz J. Molecular pathology of expanded polyalanine tract mutations in the aristaless-related homeobox gene. Genomics 2007;90;59-71.

Slager RE, Newton TL, Vlangos CN, Finucane B, Elsea SH. Mutations in *RAI1* associated with smith-magenis syndrome. Nat Genet 2003;33;466-8.

Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S et al. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. Am J Hum Genet 2005;77;709-26.

Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W et al. Genome-wide association study of bipolar disorder in european american and african american individuals. Mol Psychiatry 2009.

Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 2004;3;Article3.

Soejima H, Nakagawachi T, Zhao W, Higashimoto K, Urano T, Matsukura S et al. Silencing of imprinted CDKN1C gene expression is associated with loss of CpG and histone H3 lysine 9 methylation at DMR-LIT1 in esophageal cancer. Oncogene 2004;23;4380-8.

Solomon NM, Ross SA, Morgan T, Belsky JL, Hol FA, Karnes PS et al. Array comparative genomic hybridisation analysis of boys with X linked hypopituitarism identifies a 3.9 mb duplicated critical region at Xq27 containing SOX3. J Med Genet 2004;41;669-78.

Stankiewicz P, Beaudet AL. Use of array CGH in the evaluation of dysmorphology, malformations, developmental delay, and idiopathic mental retardation. Curr Opin Genet Dev 2007;17;182-92.

Stark KL, Xu B, Bagchi A, Lai WS, Liu H, Hsu R et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat Genet 2008;40;751-60.

Stepp ML, Cason AL, Finnis M, Mangelsdorf M, Holinski-Feder E, Macgregor D et al. XLMR in MRX families 29, 32, 33 and 38 results from the dup24 mutation in the ARX (aristaless related homeobox) gene. BMC Med Genet 2005;6;16.

Stevenson RE, Abidi F, Schwartz CE, Lubs HA, Holmes LB. Holmes-gang syndrome is allelic with XLMR-hypotonic face syndrome. Am J Med Genet 2000;94;383-5.

Stevenson RE, Schwartz CE, Schroer RJ. X-linked mental retardation. New York: Oxford University Press; 2000.

Stromme P, Mangelsdorf ME, Scheffer IE, Gecz J. Infantile spasms, dystonia, and other X-linked phenotypes caused by mutations in aristaless related homeobox gene, *ARX*. Brain Dev 2002a;24;266-8.

Stromme P, Mangelsdorf ME, Shaw MA, Lower KM, Lewis SM, Bruyere H et al. Mutations in the human ortholog of aristaless cause X-linked mental retardation and epilepsy. Nat Genet 2002b;30;441-5.

Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D et al. DNA methylation represses *FMR-1* transcription in fragile X syndrome. Hum Mol Genet 1992;1;397-400.

Swank MW, Sweatt JD. Increased histone acetyltransferase and lysine acetyltransferase activity and biphasic activation of the ERK/RSK cascade in insular cortex during novel taste learning. J Neurosci 2001;21;3383-91.

Szczaluba K, Nawara M, Poirier K, Pilch J, Gajdulewicz M, Spodar K et al. Genotype-phenotype associations for *ARX* gene duplication in X-linked mental retardation. Neurology 2006;67;2073-5.

Tabolacci E, Pietrobono R, Moscato U, Oostra BA, Chiurazzi P, Neri G. Differential epigenetic modifications in the *FMR1* gene of the fragile X syndrome after reactivating pharmacological treatments. Eur J Hum Genet 2005;13;641-8.

Tahiliani M, Mei P, Fang R, Leonor T, Rutenberg M, Shimizu F et al. The histone H3K4 demethylase SMCX links REST target genes to X-linked mental retardation. Nature 2007;447;601-5.

Tang J, Wu S, Liu H, Stratt R, Barak OG, Shiekhattar R et al. A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. J Biol Chem 2004;279;20369-77.

Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. Nat Genet 2009.

Terwilliger JD, Ott JO. Handbook of human genetic linkage. : John Hopkins University Press; 1994.

Thienpont B, de Ravel T, Van Esch H, Van Schoubroeck D, Moerman P, Vermeesch JR et al. Partial duplications of the *ATRX* gene cause the ATR-X syndrome. Eur J Hum Genet 2007;15;1094-7.

Troester MM, Trachtenberg T, Narayanan V. A novel mutation of the *ARX* gene in a male with nonsyndromic mental retardation. J Child Neurol 2007;22;744-8.

Tucker B, Richards RI, Lardelli M. Contribution of mGluR and Fmr1 functional pathways to neurite morphogenesis, craniofacial development and fragile X syndrome. Hum Mol Genet 2006;15;3446-58.

Turner G, Partington M, Kerr B, Mangelsdorf M, Gecz J. Variable expression of mental retardation, autism, seizures, and dystonic hand movements in two families with an identical ARX gene mutation. Am J Med Genet 2002;112;405-11.

Tyson C, Harvard C, Locker R, Friedman JM, Langlois S, Lewis ME et al. Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. Am J Med Genet A 2005;139;173-85.

Tzschach A, Lenzner S, Moser B, Reinhardt R, Chelly J, Fryns JP et al. Novel *JARID1C*/*SMCX* mutations in patients with X-linked mental retardation. Hum Mutat 2006;27;389.

Urban M, Chersich MF, Fourie LA, Chetty C, Olivier L, Viljoen D. Fetal alcohol syndrome among grade 1 schoolchildren in northern cape province: Prevalence and risk factors. S Afr Med J 2008;98;877-82.

Uyanik G, Aigner L, Martin P, Gross C, Neumann D, Marschner-Schafer H et al. *ARX* mutations in X-linked lissencephaly with abnormal genitalia. Neurology 2003;61;232-5.

Vallian S, Chin KV, Chang KS. The promyelocytic leukemia protein interacts with Sp1 and inhibits its transactivation of the epidermal growth factor receptor promoter. Mol Cell Biol 1998;18;7147-56.

van Bon BW, Mefford HC, Menten B, Koolen DA, Sharp AJ, Nillesen WM et al. Further delineation of the 15q13 microdeletion and duplication syndromes: A clinical spectrum varying from non-pathogenic to a severe outcome. J Med Genet 2009;46;511-23.

Van Esch H, Bauters M, Ignatius J, Jansen M, Raynaud M, Hollanders K et al. Duplication of the *MECP2* region is a frequent cause of severe mental retardation and progressive neurological symptoms in males. Am J Hum Genet 2005;77;442-53.

Van Esch H, Poirier K, de Zegher F, Holvoet M, Bienvenu T, Chelly J et al. *ARX* mutation in a boy with transsphenoidal encephalocele and hypopituitarism. Clin Genet 2004;65;503-5.

Vasconcelos MM. [Mental retardation]. J Pediatr (Rio J) 2004;80;S71-82.

Vega VB, Cheung E, Palanisamy N, Sung WK. Inherent signals in sequencing-based chromatin-ImmunoPrecipitation control libraries. PLoS One 2009;4;e5241.

Veltman JA, de Vries BB. Diagnostic genome profiling: Unbiased whole genome or targeted analysis?. J Mol Diagn 2006;8;534,7; discussion 537-9.

Veltman JA, de Vries BB. Whole-genome array comparative genome hybridization: The preferred diagnostic choice in postnatal clinical cytogenetics. J Mol Diagn 2007;9;277.

Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A et al. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell 1991;65;905-14.

Villard L, Bonino MC, Abidi F, Ragusa A, Belougne J, Lossi AM et al. Evaluation of a mutation screening strategy for sporadic cases of ATR-X syndrome. J Med Genet 1999;36;183-6.

Villard L, Fontes M, Ades LC, Gecz J. Identification of a mutation in the *XNP/ATR-X* gene in a family reported as smith-fineman-myers syndrome. Am J Med Genet 2000;91;83-5.

Villard L, Gecz J, Mattei JF, Fontes M, Saugier-Veber P, Munnich A et al. *XNP* mutation in a large family with juberg-marsidi syndrome. Nat Genet 1996;12;359-60.

Villard L, Kpebe A, Cardoso C, Chelly PJ, Tardieu PM, Fontes M. Two affected boys in a rett syndrome family: Clinical and molecular findings. Neurology 2000;55;1188-93.

Villard L, Lossi AM, Cardoso C, Proud V, Chiaroni P, Colleaux L et al. Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase. Genomics 1997;43;149-55.

Villard L, Toutain A, Lossi AM, Gecz J, Houdayer C, Moraine C et al. Splicing mutation in the ATR-X gene can lead to a dysmorphic mental retardation phenotype without alpha-thalassemia. Am J Hum Genet 1996;58;499-505.

Vissers LE, de Vries BB, Osoegawa K, Janssen IM, Feuth T, Choy CO et al. Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. Am J Hum Genet 2003;73;1261-70.

Wada T, Sakakibara M, Fukushima Y, Saitoh S. A novel splicing mutation of the *ATRX* gene in ATR-X syndrome. Brain Dev 2006;28;322-5.

Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D et al. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet 2007;81;768-79.

Wallerstein R, Sugalski R, Cohn L, Jawetz R, Friez M. Expansion of the *ARX* spectrum. Clin Neurol Neurosurg 2008;110;631-4.

Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 2008;320;539-43.

Watson P, Black G, Ramsden S, Barrow M, Super M, Kerr B et al. Angelman syndrome phenotype associated with mutations in *MECP2*, a gene encoding a methyl CpG binding protein. J Med Genet 2001;38;224-8.

Watt F, Molloy PL. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev 1988;2;1136-43.

Weatherall DJ, Higgs DR, Bunch C, Old JM, Hunt DM, Pressley L et al. Hemoglobin H disease and mental retardation: A new syndrome or a remarkable coincidence?. N Engl J Med 1981;305;607-12.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005;37;853-62.

Wieland I, Sabathil J, Ostendorf A, Rittinger O, Ropke A, Winnepenninckx B et al. A missense mutation in the coiled-coil motif of the HP1-interacting domain of ATR-X in a family with X-linked mental retardation. Neurogenetics 2005;6;45-7.

Wilkinson LS, Davies W, Isles AR. Genomic imprinting effects on brain development and function. Nat Rev Neurosci 2007;8;832-43.

Wilson IM, Davies JJ, Weber M, Brown CJ, Alvarez CE, MacAulay C et al. Epigenomics: Mapping the methylome. Cell Cycle 2006;5;155-8.

Wohlrab G, Uyanik G, Gross C, Hehr U, Winkler J, Schmitt B et al. Familial west syndrome and dystonia caused by an aristaless related homeobox gene mutation. Eur J Pediatr 2005;164;326-8.

Woods KS, Cundall M, Turton J, Rizotti K, Mehta A, Palmer R et al. Over- and underdosage of *SOX3* is associated with infundibular hypoplasia and hypopituitarism. Am J Hum Genet 2005;76;833-49.

Wu JC, Santi DV. On the mechanism and inhibition of DNA cytosine methyltransferases. Prog Clin Biol Res 1985;198;119-29.

Wu Y, Arai AC, Rumbaugh G, Srivastava AK, Turner G, Hayashi T et al. Mutations in ionotropic AMPA receptor 3 alter channel properties and are associated with moderate cognitive impairment in humans. Proc Natl Acad Sci U S A 2007;104;18163-8.

Xie S, Wang Z, Okano M, Nogami M, Li Y, He WW et al. Cloning, expression and chromosome locations of the human *DNMT3* gene family. Gene 1999;236;87-95.

Xue Y, Gibbons R, Yan Z, Yang D, McDowell TL, Sechi S et al. The ATR-X syndrome protein forms a chromatin-remodeling complex with daxx and localizes in promyelocytic leukemia nuclear bodies. Proc Natl Acad Sci U S A 2003;100;10635-40.

Yagi H, Furutani Y, Hamada H, Sasaki T, Asakawa S, Minoshima S et al. Role of *TBX1* in human del22q11.2 syndrome. Lancet 2003;362;1366-73.

Yamamoto G, Nannya Y, Kato M, Sanada M, Levine RL, Kawamata N et al. Highly sensitive method for genomewide detection of allelic composition in nonpaired, primary tumor specimens by use of affymetrix single-nucleotide-polymorphism genotyping microarrays. Am J Hum Genet 2007;81;114-26.

Yan QJ, Rammal M, Tranfaglia M, Bauchwitz RP. Suppression of two major fragile X syndrome mouse model phenotypes by the mGluR5 antagonist MPEP. Neuropharmacology 2005;49;1053-66.

Yasui DH, Peddada S, Bieda MC, Vallero RO, Hogart A, Nagarajan RP et al. Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc Natl Acad Sci U S A 2007;104;19416-21.

Yntema HG, Poppelaars FA, Derksen E, Oudakker AR, van Roosmalen T, Jacobs A et al. Expanding phenotype of *XNP* mutations: Mild to moderate mental retardation. Am J Med Genet 2002;110;243-7.

Zahir F, Firth HV, Baross A, Delaney AD, Eydoux P, Gibson WT et al. Novel deletions of 14q11.2 associated with developmental delay, cognitive impairment and similar minor anomalies in three children. J Med Genet 2007;44;556-61.

Zahir F, Friedman JM. The impact of array genomic hybridization on mental retardation research: A review of current technologies and their clinical utility. Clin Genet 2007;72;271-87.

Zechner U, Kohlschmidt N, Rittner G, Damatova N, Beyer V, Haaf T et al. Epimutation at human chromosome 14q32.2 in a boy with a upd(14)mat-like clinical phenotype. Clin Genet 2009;75;251-8.

Zechner U, Wilda M, Kehrer-Sawatzki H, Vogel W, Fundele R, Hameister H. A high density of X-linked genes for general cognitive ability: A run-away process shaping human evolution?. Trends Genet 2001;17;697-701.

Zemni R, Bienvenu T, Vinet MC, Sefiani A, Carrie A, Billuart P et al. A new gene involved in X-linked mental retardation identified by analysis of an X;2 balanced translocation. Nat Genet 2000;24;167-70.

Zendman AJ, Zschocke J, van Kraats AA, de Wit NJ, Kurpisz M, Weidle UH et al. The human *SPANX* multigene family: Genomic organization, alignment and expression in male germ cells and tumor cell lines. Gene 2003;309;125-33.

Zhong S, Salomoni P, Pandolfi PP. The transcriptional role of PML and the nuclear body. Nat Cell Biol 2000;2;E85-90.

Zollino M, Murdolo M, Marangi G, Pecile V, Galasso C, Mazzanti L et al. On the nosology and pathogenesis of wolf-hirschhorn syndrome: Genotype-phenotype correlation analysis of 80 patients and literature review. Am J Med Genet C Semin Med Genet 2008;148C;257-69.